UNIVERSITY OF LONDON THESIS

Degree phd Year 2007 Name of Author MICHAEL HARRY FIELD

IN VITRO AND *IN SILICO* EXAMINATION OF THE ALTERNATIVE RESPIRATORY NADH DEHYDROGENASE FAMILY IN *ARABIDOPSIS THALIANA* AND OTHER SPECIES.

A THESIS SUBMITTED FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

MICHAEL FIELD

DEPARTMENT OF BIOLOGY

UNIVERSITY COLLEGE LONDON

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

The respiratory electron transport chain (ETC) in plants, fungi and many bacteria incorporates additional enzymes not present in mammalian counterparts. The alternative NADH dehydrogenase (NDH2) performs the same enzymatic function as Complex I - the oxidation of NADH and reduction of membrane-bound quinone - but without the concomitant translocation of hydrogen ions across the ETC membrane, and thus does not contribute directly to the proton-motive potential. Three genes in *Saccharomyces cerevisiae* and as many as seven putative genes have been identified in *Arabidopsis thaliana* encoding NDH2 enzymes yet their structure, enzymatic kinetics and the benefits they confer on the organisms remain speculative.

This thesis reports the investigation of these enzymes using a combination of computational and biochemical techniques. cDNA clones were produced from *Arabidopsis* leaf tissue and subcloned into an *E. coli* expression vector from which a high yield of protein was obtained, but this aggregated as inclusion bodies. Refolding assays using dialysis and injection diffusion techniques failed to obtain active protein. NDH2 gene sequences from all species were reviewed in detail both to determine expression patterns and to identify gene-specific motifs. The latter were used in a novel homology modelling process which has determined key structural elements, including residues specific for NADH or NADPH and the fact that these enzymes have been incorrectly annotated as "NADH disulphide oxidoreductases". The modelling analysis yielded other enzyme properties pertinent to biochemical expression, such as the absence of disulphide bridges in the tertiary structure.

Acknowledgements

I am grateful to my supervisor, Prof. Peter Rich, for the time and effort he has contributed to this research and for his corrections to this thesis.

I wish to thank also the other members of the laboratory, Dr Masayo Iwaki and Vasanti Amin for technical assistance, Dr Wendy Fairclough for instruction in Molecular Biology techniques, and my fellow student, Doug Marshall, for maintaining my sanity.

In loving memory of my mother who died 14 July 2006

This research has been generously funded by a grant from BBSRC

Table of Contents

Chapter 1. Introduction

Chapter 2. Materials and Methods

Chapter 3. NDH2 Sequence Analysis

Chapter 4. Cloning and Expression of *Arabidopsis* nda1

Chapter 5. *In silico* Investigation into NDH2 Structural Features

Chapter 6. Discussion

List of Figures

List of Tables

Glossary of terms

µg	Microgram
µl	Microlitre
µm	Micrometer
µM	Micromolar
µs	Microseconds
ADP	Adenosine diphosphate
APS	Ammonium persulphate
ATP	Adenosine triphosphate0.
bp	Base pair
BSA	Bovine serum albumin
C-terminus	Carboxyl terminus
CaMV	Cauliflower mosaic virus
CoM	Coenzyme M
cyt *c*	Cytochrome *c*
Da	Dalton
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytidine triphosphate
DEPC	Diethylpyrocarbonate
dH_2O	Distilled water
ddH_2O	Double-distilled water
dGTP	Deoxyguanosine triphosphate
DM	Dodecyl maltoside
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
DPI	Diphenyleneiodonium
DTT	Dithiothreitol
dTTP	Deoxythymidine triphosphate
EST	Expressed sequence tag
ETC	Electron transport chain
FeCN	Ferricyanide (potassium)
Fe-S	Iron-sulphur (cluster)
GITC	Guanidine isothiocyanate

GFP	Green fluorescent protein
GSH	Glutathione (reduced)
GSSG	Oxidised (disulphide) Glutathione
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl-b-d-thiogalactopyranoside
kDa	Kilodalton
KPi	Potassium phosphate buffer (20 mM)
LB	Luria-Bertani medium
M	Molar
MAE	Maleic acid
MCS	Multiple cloning site
mg	Milligram
ml	Millilitre
mm	Millimetre
mM	Millimolar
ms	Millisecond
N-terminus	Amino terminus
NAD^+	Nicotinamide adenine dinucleotide (oxidised)
NADH	Nicotinamide adenine dinucleotide (reduced)
$NADP^+$	Nicotinamide adenine dinucleotide phosphate (oxidised)
NADPH	Nicotinamide adenine dinucleotide phosphate (reduced)
ng	Nanogram
NR	Non-redundant
ns	Nanosecond
OG	Octylglucoside
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PEG	Polyethylene glycol
PCR	Polymerase chain reaction
PDB	Protein DataBank (structural models)
psi	Pounds per square inch
Q_1	Ubiquinone-5
Q_2	Ubiquinone-10
ROS	Reactive Oxygen Species

RT	Room temperature
RT-PCR	Reverse transcription-PCR
SDS	Sodium dodecyl sulphate
SHAM	Salicylhydroxamic acid
Spdbv	Swiss PDB viewer (graphics software)
TAE	Tris-acetate-EDTA
TCA	Tricarboxylic acid
TEMED	N,N,N',N'-Tetramethylethylenediamine
Tris	Tris(hydroxymethyl)aminomethane
UCP	Uncoupling Protein
UV	Ultraviolet
v/v	Volume for volume
w/v	Weight for volume
X-gal	5-bromo-4-chloro-3-indolyl b-D-galactopyranoside

Chapter 1. Introduction

1.1 Chemiosmosis and the mammalian Electron Transport Chain

A pivotal contribution to the elucidation of respiratory bioenergetics was the discovery that ADP was recycled to ATP *not* through substrate level phosphorylation but by a molecular turbine, ATP synthase. Power for this is provided by the combined pH and electric potential that exists across the membrane in which the synthase is located. This 'chemiosmotic' force arises from the relative concentration of hydrogen ions on each side of the membrane and was first proposed by Peter Mitchell (Mitchell, 1961), forming the basis of his 'chemiosmotic theory'.

The chemiosmotic potential arises from the net translocation of protons across the membrane performed by components of the Electron Transport Chain (ETC): four multiunit redox-active complexes electrically connected by membrane (ubiquinone) and soluble (cytochrome *c*) electron carriers (fig 1.1). Proton translocations occurs at Complexes I, III and IV. Electrons enter the ETC from the oxidation of NADH (Complex I) and succinate (Complex II), both products of the tricarboxylic acid cycle, and are transferred to quinone (coenzyme Q) intermediaries within the membrane. The quinone 'pool' thus comprises molecules in both oxidised and reduced (ubiquinol)

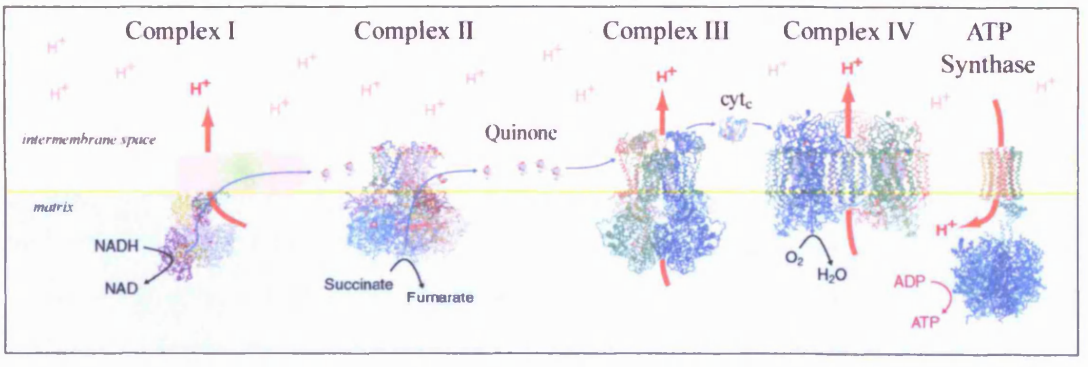


Fig 1.1 Principal components of mammalian respiration.
Complex I: NADH dehydrogenase/ Ubiquinone reductase; Complex II: Succinate dehydrogenase; Complex III: Cytochrome bc_1 complex; Complex IV: cytochrome c oxidase; cyt_c: cytochrome c; H^+: proton.
Red arrows: proton net translocation; blue arrows: electron pathways.
Adapted from Rich (2003); Complex I hydrophilic domain from (bacterial)*Thermus thermophilus* (PDB ID: 2FUG) included for schematic illustrative purposes. Complex I membrane domain not yet resolved.

states[1]. The energy released by the transfer of electrons from quinol to cytochrome *c* (cyt *c*) via redox sites within Complex III fuels proton translocation, as does oxidation of cyt *c* via Complex IV redox sites to the final electron acceptor, molecular oxygen.

This elaborate apparatus has subsequently been found in virtually all eukaryotic mitochondria, and a related chemiosmotic mechanism is also responsible for ADP phosphorylation in plant and algal chloroplasts, as well as in the membranes of photosynthetic bacteria. Yet more remarkable is that similar respiratory systems have been discovered in many bacterial species where, lacking a dedicated respiratory organelle, the periplasmic membrane serves as the chemiosmotic barrier. Bacterial ETC components are less complex (ie. fewer subunits) and there is some variation in the implementation of the chain. Complexes III and IV are replaced in *Escherichia coli*, for example, by a single enzyme that oxidises quinol using molecular oxygen as the final electron acceptor.

Since elucidation of the core mechanism the ETC has been subjected to intense analysis and atomic-level crystal structures have been resolved for Complexes II, III and IV and for the hydrophilic domain of bacterial Complex I (see 1.2), although the structure of the membrane domain of Complex I remains elusive. Inhibitors have also been identified that target the major ETC components (Complex I: rotenone; Complex III: antimycin A; Complex IV: potassium cyanide; among others). These have provided conclusive evidence for a single electron pathway in mammalian mitochondria.

[1] As is the case with $NADH/NAD^+$, oxidised and reduced states differ by two electrons; a highly reactive single electron anion intermediary, semiquinone, can also be formed.

1.2 Structure of Complex I

Mitochondrial Complex I (NADH dehydrogenase/ubiquinone reductase) transfers electrons from matrix NADH to membrane-bound ubiquinone via its cofactors: FMN and seven or eight redox-active Fe-S centres. The energy released by the flow of electrons is used to fuel the translocation of protons from the matrix to the intermembrane space, at a stoichiometry believed to be four protons per electron pair.

Although mitochondrial Complex I has not been resolved to the atomic level, a coarser level structure for the bovine enzyme, determined at 22Å by electron cryo-microscopy (Grigorieff, 1998; see also Carroll *et al.*, 2003), revealed the classic 'L' shape that has been found in other eukaryotes, such as *Neurospora crassa* (Guénebaut *et al.*, 1997) and (in greater detail for) *Yarrowia lipolytica* (Radermacher *et al.*, 2006). The 'L' shape may be seen as being comprised of two subunits: a membrane-embedded (hydrophobic) domain, presumed to be responsible for proton translocation, and a peripheral (hydrophilic) domain that extends into the matrix. The latter contains most (or all) of the redox centres of the electron pathway: seven or eight Fe-S clusters and a noncovalently-bound FMN that accepts the initial pair of electrons from NADH. The genes for mitochondrial Complex I are encoded in both the nuclear and mitochondrial genomes.

Bacterial Complex I is a far smaller, ~530kDa compared to ~890kDa for the bovine enzyme (Grigorieff, 1998), and is comprised of typically 14 subunits (13[1] to15, see below) rather than the 45 (at least) of the bovine enzyme (Carroll *et al.*, 2006). However, all 14 peptides have unambiguous homologues in the mitochondrial Complex I gene complement, there is a similar complement of redox sites (FMN plus eight or nine Fe-S clusters), and the topology is similarly 'L' shaped with membrane and peripheral domains. It is thus reasonable to propose that the bacterial enzyme represents a 'minimal' implementation of the mitochondrial enzyme (Guénebaut *et al.*, 1998).

[1] In *E. coli,* for example, two genes are fused (NuoC and NuoD), so this organisms has only 13 Complex I peptides.

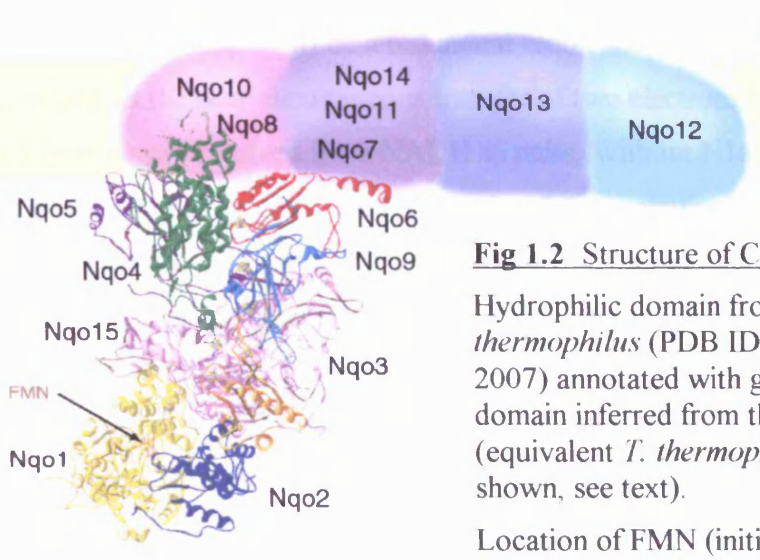


Fig 1.2 Structure of Complex I

Hydrophilic domain from *Thermus thermophilus* (PDB ID: 2FUG, after Sazanov, 2007) annotated with gene names; membrane domain inferred from that of *Escherichia coli* (equivalent *T. thermophilus* gene names shown, see text).

Location of FMN (initial electron acceptor) shown by arrow.

Although no high-resolution data have been obtained for intact Complex I, the peripheral domain from *Thermus thermophilus* has been resolved at 3.3Å (Sazanov & Hinchliffe, 2006), and projection maps from 2D crystals of the membrane domain from *Escherichia coli* have been produced at 8Å (Baranova *et al.*, 2006). From this it is possible to infer the structure of the complete *T. thermophilus* enzyme (fig 1.2).

Attempts to purify intact Complex I from *T. thermophilus* were unsuccessful as the peripheral domain dissociates easily (Hinchliffe *et al.*, 2006). However crystals formed from the purified peripheral domain afforded the determination of the structure at 3.3Å with 93% of the residues (as predicted from sequence data) successfully resolved. (Some terminal and loop regions are omitted.) The structure is described as having a 'Y' shape (more apparent in fig 1.1), consistent with previous low resolution data. The model includes seven conserved peptides of the peripheral domain (Nqo1, 2, 3, 4, 5, 6, 9), the non-covalently bound FMN and nine Fe-S clusters[1] (fig 1.3). The structure also shows the presence of a 15th peptide, Nqo15, which is not present in the genomes of most bacteria. On the basis of sequence data, this gene has only been identified in related extremophiles (*Deinococcus radiodurans* and *Deinococcus geothermalis*) so it may serve to stabilise the enzyme structure.

[1] This group had previously identified the *T. thermophilus* Fe-S clusters by X-ray crystallography (Hinchliffe & Sazanov, 2005).

An unambiguous electron pathway was revealed between the redox cofactors N3 and N2, the latter being the terminus of the chain from which electrons are transferred to quinone. N1a might appear to be a redundant redox site, but it is proposed that its presence allows the near-simultaneous transfer of two electrons from FMN (to N3 and N1a). Electrons are received from NADH as pairs. Without N1a the FMN molecule would become (albeit briefly) a highly reactive flavosemiquinone, which could give rise to ROS by reaction with oxygen. N1a is too far from N3 to permit direct transfer (22.3Å where 14Å is considered the maximum) so its electron must return to N3 via FMN. As N1a is conserved the protection offered by this mechanism is clearly highly beneficial. By contrast, N7 is too far from its nearest neighbour (24.2Å) to be able to contribute to electron transfer. It is also found in *E. coli* (and some other bacteria), and may serve some structural purpose, or it may, it is proposed, simply be an evolutionary relic.

Projection maps at a resolution of 8Å were obtained from stained 2D crystals of the membrane domain from *E. coli* Complex I. These revealed a curved shape consistent with models for other species including eukaryotic (yeast and bovine). The attachment to the membrane domain (fig 1.2) was inferred from similar analysis (by comparing images derived from intact enzyme with those from membrane domain preparations)

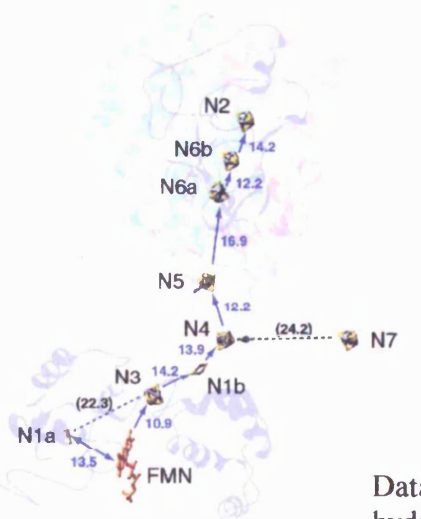


Fig 1.3 Complex I Electron Pathway

Electron pairs are transferred to the first cofactor, FMN, from soluble NADH. Single electrons then flow via N3 (either directly or via N1a) and ultimately to N2 from where they are transferred to a quinone molecule. The distance of N7 from its nearest neighbour (N4) suggests that it an evolutionary relic, not actively involved in the transfer of electrons.

Data from PDB ID 2FUG (*Thermus thermophilus* hydrophilic domain); cofactor separations shown in blue (after Sazanov, 2007).

using *Neurospora crassa*. The similarity of images at this resolution implies that, as for the peripheral domain, the bacterial membrane domain represents a 'minimal' implementation of its mitochondrial counterpart.

Tentative assignments of peptide locations within the model are made, based in part on probable transmembrane helices in the images compared with predicted secondary structures. This identifies NuoL (=Nqo12 in *T. thermophilus*, fig 1.2) at the distal end (with respect to the peripheral domain attachment), adjacent to NuoM (Nqo13). Crosslinking experiments had suggested that NuoA (Nqo7) interacts with NuoB and NuoD (Nqo6 and Nqo4) of the peripheral domain, and NuoH (Nqo8) has been implicated in quinone interaction (Yagi & Matsuno-Yagi, 2003) so these are likely to be proximal to the peripheral domain.

Evidently, proton translocation occurs within the membrane domain, though the exact implementation and the mechanism by which it is coupled to electron flow remains unresolved. Two coupling mechanisms have previously been postulated: direct (redox-driven), and indirect (long-range conformational changes resulting from the redox state changes of the clusters). NuoN (Nqo14), NuoL and NuoM share homology to each other (and to putative antiporter genes) and are proposed as proton channels. Given the stoichiometry of Complex I (four proton translocations per electron pair from NADH to quinone), the authors propose that these subunits might translocate three protons indirectly (extremely so in the case of distal NuoL) while the fourth electron may be translocated directly. It should be noted though that menaquinone, the electron acceptor in *T. thermophilus*, has a midpoint redox potential of approximately -80mV while bovine stoichiometric assays, which determined the 4 proton/2electron ratio, used ubiquinone (midpoint potential approx. +110mV). The difference in the NADH/Q redox couple might result in a different stoichiometry in *T. thermophilus*.

1.3 Additional enzymes in non-mammalian systems

An unexpected early discovery was that externally introduced NADH and NADPH is oxidised by isolated plant mitochondria (unlike mammalian preparations), although it

was known that Complex I accepts NADH from the matrix alone, and redox shuttles (eg. glycerol-3-phosphate and ethanol-acetaldehyde shuttles) are inoperative in the absence of their cytosolic components (Bonner & Voss, 1961).

Plant cells are not alone in possessing an alternative pathway for NADH dehydrogenase activity coupled to the ETC. Investigation into a number of species using rotenone inhibition of Complex I, deletion mutations, and biochemical analysis of mitochondrial preparations has revealed a family of "alternative NADH dehydrogenases" (NDH2s) associated with both inner and outer faces of the mitochondrial cristae (and the periplasmic membrane of bacteria). These are reviewed in detail in the following sections. Indeed, EPR analysis and rotenone insensitivity have long suggested that Complex I is not expressed (almost uniquely) in *Saccharomyces cerevisiae* (baker's yeast); and more recently completion of the genome sequencing project has confirmed that functional genes are not encoded. Nevertheless *S. cerevisiae* is capable of respiratory growth[1].

NDH2s provide an 'alternative' route for entry of electrons into the ECT. Investigation into the terminal oxidase (eg. Complex IV) has revealed that in many organisms, other than animals, there exists also an alternative pathway for the exit of electrons, generally to molecular oxygen although a variety of acceptors is found in bacterial species. This function is performed by one or more[2], nuclear-encoded enzymes, 'AOX', the alternative oxidase, which transfers electrons from quinone to oxygen, bypassing

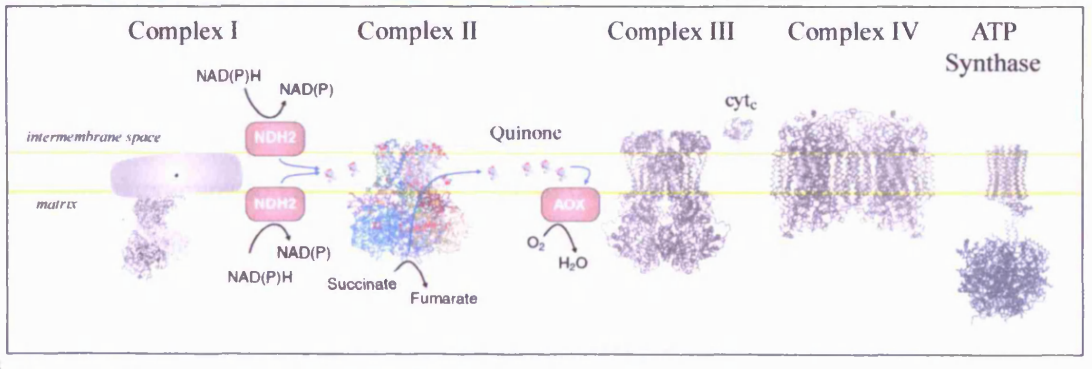


Fig 1.4 Alternative respiratory ETC enzymes.
NDH2: alternative NADH Dehydrogenase; AOX: Alternative Oxidase. The electron route shown (blue arrows) bypasses all proton-translocation respiratory enzymes. (Compare with fig 1.1)

[1] *Saccharomyces carlsbergensis* too does not encode a functional Complex I. This may be a consequence of hundreds of years of selection for fermentation performance.
[2] five *aox* genes have been identified in *Arabidopsis* (Clifton et al. 2006).

Complex III, Complex IV and cyt *c* (review: Juszczuk & Rychter, 2003; McDonald & Sieger, 2002). These enzymes' activity is readily distinguished from Complex IV as it is insensitive to cyanide but inhibited by salicylhydroxamic acid (SHAM) and has been found in plant and fungi species - though not in *S. cerevisiae*. In trypanosomes (where it is termed the Terminal Alternative Oxidase, 'TAO') it is being investigated in some detail (Chaudhuri *et al.*, 2005; Chaudhuri *et al.*, 2006), as it is a potential target for therapeutic intervention. Together NDH2 and AOX offer an electron route that potentially bypasses proton translocating enzymes entirely (fig 1.4), each converting the released energy to heat rather than ADP phosphorylation. In the case of the external-facing NDH2s a mechanism is additionally provided for the direct re-oxidation of cytosolic NAD(P)H.

However it would be facile to conclude that these additional components serve solely to permit ETC operation while circumventing ADP phosphorylation, as an alternative mechanism exists for this. The enzyme in question, closely associated with the regulation of the chemiosmotic machinery, is the Uncoupling Protein (UCP) expressed particularly in mammalian adipose tissue as well as in NDH2/AOX-expressing species. This exploits a fatty acid recycling shuttle positioned across the inner membrane that effectively 'short circuits' the chemiosmotic potential, converting this latent energy into heat.

As described here the ETC and associated chemiosmotic components might appear to constitute a discrete unit, functionally and (in the case of mitochondria) physically removed from other cellular metabolic pathways. But it is essential to consider the interactions these enzymes and substrates have with their environment - both in biochemical examination and in the drawing of conclusions as to function. NADH, for example, is a substrate for many enzymes and also can exchange its redox state with NADP using the transhydrogenases (Bykova *et al.*, 1999); it is conveyed (more accurately, its redox state) by numerous additional shuttles across the mitochondrial membrane (Bakker *et al.*, 2000); and maintenance of its net redox state ($NADH:NAD^+$ ratio) plays a crucial cellular regulatory as well as metabolic role (Lin & Guarente, 2003). Furthermore, ATP synthase is not the sole beneficiary of the chemiosmotic potential - it's energy may be used also to power other membrane-resident enzymes.

1.4 The alternative NADH dehydrogenase

Complex I and its bacterial homologues are generally classified as 'Type I' respiratory NADH dehydrogenases, while NDH2 enzymes are termed 'Type II' (review: Melo *et al.*, 2004). They perform a similar enzymatic function - the transfer of two electrons from a soluble NADH molecule to a membrane-bound quinone - but the disparity in their structural composition is considerable. Bovine Complex I (Carroll *et al.*, 2003) has a mass of 980kDa and is composed of at least 46 protein elements, one FMN molecule and at least six iron-sulphur centres (both Fe_4S_4 and Fe_2S_2 types)[1]. Further compounding the task of stoichiometric expression, genes for the peptides are encoded in both the mitochondrial (seven genes) and nuclear genomes. In contrast NDH2 is typically just 60kDa, encoded in a single nuclear gene and contains a single prosthetic group, almost universally FAD.

The only known functional distinction between the two dehydrogenases is that in Type II enzymes the energy released by the redox changes is 'lost' as heat, whereas in Type I enzymes it is used to translocate hydrogen ions (generally believed to be four per NADH). Clearly this sophisticated mechanism requires the complex assembly of multiple subunits and redox sites. The maintenance of this substantial genetic investment (particularly in the case of the bacteria) pays testimony to the selection benefits afforded by oxidative phosphorylation.

NDH2s are 'alternative' only in the sense of historical discovery, and in no way constitute some manner of 'back-up system'. Species exist which express solely Complex I (such as the bacterium *Paracoccus denitrificans*) while, as has been stated, *S. cerevisiae* copes with just NDH2s. Plant and most fungal species examined to date express both a Complex I and a repertoire of NDH2s - as many as seven for *Arabidopsis*[2] (Michalecka *et al.*, 2003) - in what must be assumed to be an adaptive, coordinated symphony. Knock-out strains targeting either type typically exhibit degraded growth as reviewed below.

---

[1] One Fe_2S_2 and six Fe_4S_4 in the recently determined bacterial *Thermus thermophilus* (Hinchliffe & Sazanov, 2005)

[2] Throughout this thesis '*Arabidopsis*' refers to *Arabidopsis thaliana*.

1.5 NDH2 research

1.5.1 Nomenclature

A wealth of information is accruing from the examination of NDH2s from species as diverse as higher plants and thermophilic bacteria, through both biochemical and genetic analysis (review: Kerscher, 2000).

Some confusion in nomenclature has arisen, perhaps inevitably, from independent research groups whose attention has focused on differing properties of these enzymes. Earlier research (notably with *S. cerevisiae*) has adopted suffixes denoting the location of the enzymes as determined by biochemical experiments: 'NDI' and 'NDE' refer to enzymes associated with the Internal and External faces of the mitochondrial inner membrane respectively. More recently (Svensson *et al.*, 2002; Michalecka *et al.*, 2003) genetic examination classifies NDH2s (whose location is not determined, at least initially) by sequence homology: **nda**, **ndb** and **ndc** in higher plants, a system which carries the additional benefit of an implied functional distinction. (**ndb** sequences incorporate a putative 'EF-hand' calcium binding domain, although no functional distinction is evident between **nda** and **ndc**.)

1.5.2 Bacterial NDH2

An NDH2 was first identified in *Escherichia coli* a quarter of a century ago (Jaworowski *et al.*, 1981) revealing a 45kDa enzyme by SDS-PAGE and characterising its turnover at 0.5 mmol NADH/min/mg enzyme. Since then this model organism has received surprisingly scant attention in NDH2 research. Indeed the first expression of its gene ('*ndh*') was for the purpose of recovering a mutant strain in *Parococcus denitrificans* (Finel, 1996). More recently the gene has been engineered to incorporate a His-tag and overexpressed in the native host (Björklöf *et al.*, 2000; and see 1.5).

Two papers from the same laboratory suggest an alternative interpretation of NDH2 cellular function (Rapisarda *et al.*, 1999; Rapisarda *et al.*, 2001). These propose a cupric

reductase role for *Escherichia coli* NDH2 and a thiolate-bound Cu(I) in the enzyme is claimed, though these observations are not shared elsewhere in the NDH2 research literature.

An investigation using purified *Corynebacterium glutamicum* NDH2 identified a pH-dependent specificity for NADH/NADPH (Matsushita *et al.*, 2001). A Triton-X solubilised membrane fraction was subjected to two column chromatographic purification steps which yielded a protein of 55kDa. Peptide sequencing of this indicated that it was the product of a putative *ndh2* gene previously identified in the genome. In enzyme assays of the purified protein it was shown that at pH4.5 it exhibits a preference for NADPH, while at physiological pH NADH is the preferred substrate. Examination of enzyme substrate preference at pH levels well beyond the physiological might seem curious, but it had already been established that NADH/NADPH specificity was a property that distinguishes some members of the NDH2 family. Bacterial NDH2s examined to date appear to be specific for NADH and do not incorporate the EF-hand motif.

Respiratory system investigation using cyanobacteria presents an interesting challenge as active ETCs reside in both cytoplasmic and thylakoid membranes, with interactions between photosynthetic and respiratory systems further compounding the analysis. Three NDH2 genes have been identified in *Synechocystis* PCC 6803 (Howitt *et al.*, 1999), and their role proposed as regulatory rather than metabolic. Two of the three genes were individually cloned into expression vectors and used to complement an *E. coli* strain deficient in both Complex I and NDH2, although only one of these proved successful in restoring respiratory growth. Individual and multiple NDH2 deletions were created in a PSI-deficient *Synechocystis* strain from which it was concluded that in all cases light tolerance was *enhanced.* It is proposed (uniquely in NDH2 publications) that NDH2 has a regulatory function monitoring the redox state of the PQ pool.

An illuminating avenue for exploration is the adaptation of organisms and their metabolic systems to extreme environmental conditions (Pereira *et al.*, 2004). *Acidanus ambivalens* is a hyperthermoacidophilic archaeon thriving at pH2.0 and 80°C. It has evolved a simplified respiratory chain which has been examined using membrane preparations and artificial liposomes containing respiratory components (Gomes *et al.*,

2001) and its NDH2 gene has been identified (Bandeiras *et al.*,2002). The ETC was determined to include a succinate dehydrogenase, modified quinone ('caldariella quinone') and a single cyanide-sensitive quinol terminal oxidase, but no Complex I (on the basis of rotenone insensitivity to NADH-O_2 oxidoreductase activity). A 47kDa NDH2, purified from membranes by progressive DM solubilisation, differs from the canonical structure in that it employs FMN rather FAD as the cofactor, and that this prosthetic unit is *covalently* bound as it does not dissociate from denatured protein. However comparison of genomic sequence data confirms an unambiguous relationship to NDH2 genes from other *archea* species. NADH dehydrogenase activity using the FeCN assay was significantly attenuated by detergent solublilisation, as is commonly reported but, interestingly, could be restored by the addition of sonicated phospholipids as had been demonstrated in the case of *E. coli* (Björklöf *et al.*, 2000).

Another, though less extreme hyperthermoacidophilic archaeon, *Sulfolobus metallicus* (optimum growth 65°C, pH2.5), expresses a 49kDa NDH2 with a covalently bound FMN cofactor as determined by ^{31}P-NMR (Bandeiras *et al.*, 2003). Enzyme assays were performed on the purified protein both with FeCN and, following addition of 1.2 mg/ml asolectin, with ubiquinone-2, revealing a turnover of ~0.2 mmol/min/mg and specificity for NADH. A slightly lower redox potential was observed (~90mV) when compared to *A. ambivalens* (above, ~70mA) both significantly lower than that of the *S. cerevisiae* NDI1 (-370mV).

Azotobacter vinelandii is atypical of bacteria in that it can readily fix nitrogen in the presence of oxygen even though its nitrogenase complex, in common with others, is susceptible to oxygen damage. An efficient ETC has been credited with the minimisation of cytosolic O_2 with NDH2 a major contributor to this process (Bertsova *et al.*, 2001). In this organism it was shown that NDH2 expression is upregulated in response to increasing cytosolic O_2 concentration.

1.5.3 Fungal NDH2

Fungi have been the subject of respiratory research, both for the 'conventional' and for the alternative pathways (review: Joseph-Horne *et al.*, 2001). Lacking a Complex I, yet still capable of respiration, *Saccharomyces cerevisiae* has proved a particularly appropriate candidate for NDH2 research. A 53kDa protein, isolated from the mitochondrial membranes by Triton-X solubilisation and chromatographic purification, was shown to oxidise NADH in the presence of either FeCN or ubiquinone, and to be more abundant when the culture was grown on glucose as the sole carbon source (de Vries & Grivell, 1988). The enzyme was shown to contain non-covalently linked FAD, had an apparent K_m of 31μM, redox potential of -370mV, was insensitive to the Complex I inhibitors, rotenone and piericidin, but *was* inhibited by flavone. Initially reported as an external-facing NDH2 it was subsequently shown to be matrix-facing (Marres *et al.*, 1991), and its (nuclear-encoded) gene was identified. An unexpected observation[1] was that growth rate on glucose was not affected in a strain engineered to disrupt the gene, suggesting a more complex interplay of metabolic pathways. This gene ('NDI1' in current nomenclature) was sequenced (de Vries *et al.*, 1992) and the cleavage site for its import signal determined by N-terminus sequencing of the mature protein. *In vitro* experiments using a mitochondrial preparation to examine the importation mechanisms suggested, curiously, that a pre-existing membrane potential was required (successful importation being determined by resistance to proteinase K treatment). Further characterisation of NDI1 has been achieved through overexpression in *E. coli* (Kitajima-Ihara & Yagi, 1998) and it has also found application in biomedical research (see 1.7).

With completion of the *S. cerevisiae* genome sequencing project, homology searching revealed two further putative *S. cerevisiae* NDH2s, examined simultaneously by independent groups (Small & McAlister-Henn, 1998; Luttik *et al.*, 1998), the latter assigning the currently recognised names: 'NDE1' and 'NDE2'. Both groups created '*Δnde*' deletion strains to examine the phenotype consequences. Luttik and co-workers demonstrated that mitochondrial preparations obtained from *Δnde1*/*Δnde2* (double mutants) failed to oxidise NADH as might be expected. However cultures of the double mutants nevertheless used respiratory pathways when grown on glucose. This implies

[1] particularly in the light of the current understanding that this the sole matrix-facing NDH2.

that additional mechanisms (mitochondrial redox shuttles) must exist to reoxidise cytosolic NADH, thus highlighting the difficulty of examining elements of a pathway in isolation. They also reported that levels of ethanol in commercially obtained NADH (0.5 mol EtOH /mol NADH) result in misleading O_2 depletion rates as ethanol is also readily metabolised via the respiratory pathway, and solved this by developing an alternative method for synthesising NADH. Single mutant strains (Δ*nde1* and Δ*nde2*) exhibited wild type growth rates when supplied with ethanol, while double mutant growth was attenuated by around 35%.

Neurospora crassa NDH2s have been examined in some detail by a group in Portugal (Videira & Duarte, 2002). The common theme to their approach has been to clone the gene in order to obtain polyclonal antibodies and then use Western blotting to identify the enzyme location (internal or external) by progressive disruption of mitochondrial preparations. NDE1 (initially also named p64) was first identified as a matrix-facing enzyme (Melo *et al.*, 1999) but subsequently as an external NDH2 (Melo *et al.*, 2001) incorporating a calcium binding domain. In this respect *N. crassa* differs from *S. cerevisiae* as none of the three NDH2 genes of the latter contain the EF-hand motif associated with calcium binding. Using mitochondrial preparations (intact and IO-SMP[1]) from NDE1-mutant and wild type strains they concluded that this enzyme has a significant preference for NADPH over NADH, a functional distinction which was becoming evident in other calcium-dependent NDH2s. A more extensive examination of NDI1 (Duarte *et al.*, 2003) also identified the mature N-terminus by Edman degradation of the purified protein. By chance this also revealed an atypical intron site (GC-AG) at this locus[2]. NDE2 was shown to be the major contributor to cytosolic NADH oxidation (Carneiro *et al.*, 2004) in a study using double (Δ*nde1*/Δ*nde2*) and triple (Δ*nde1*/Δ*nde2*/Δ*ndi1*) mutant strains. They reported also that transcription of both NDE2 and NDI1 (but not NDE1 or Complex I) is significantly down-regulated during the late exponential growth phase. An unexpected result was the inability to obtain crosses from NDE2 and Complex I deficient mutants.

[1] Inside Out Sub-Mitochondrial Particles.
[2] and, incidentally, conflicting with the Genbank deposited predicted protein

Yarrowia lipolytica, untypically among fungi, encodes a single NDH2[1]. This conclusion was inferred from PCR amplification of genomic DNA using primers targeted at conserved NDH2 sequence motifs (Kerscher *et al.*, 1999). This single NDH2 enzyme is external-facing so Complex I mutations are not surprisingly lethal, though NDH2 deletion mutations in this obligate aerobe reveal that growth on either glucose or acetate is unaffected. By substituting the *nuam* (Complex I subunit) targeting signal for the endogenous NDH2 N-terminal, the enzyme was relocated as a matrix-facing NDH2 (Kerscher *et al.*, 2001). This was successfully employed to rescue respiratory growth in a Complex I deficient strain.

1.5.4 Higher plant NDH2

Plant species have been the focus for significant research into NDH2s (Affourtit *et al.*, 2001; Møller, 2002; Rasmusson *et al.*, 2004). It has been known for some time that plant mitochondrial preparations, unlike those from mammals, oxidise NADH in the presence of the Complex I inhibitor, rotenone, albeit at an attenuated rate (Ikuma & Bonner, 1967). Subsequent investigations had identified two separate NAD(P)H dehydrogenase activities located to the matrix-facing side of the inner mitochondrial membrane (Møller & Palmer, 1982) only one of which was rotenone sensitive and enzymatically active using deamino-NAD(P)H. Two distinct external NAD(P)H dehdrogenase activities had also been identified using intact artichoke mitochondrial preparations and the activity of the latter had been shown to be dependent on calcium (Møller *et al.*, 1981). Two distinct matrix-facing NDH2s (in addition to Complex I) were identified in potato (Melo *et al.*, 1996), one of which is specific for NADH, calcium-independent and insensitive to DPI, while the other is inhibited by DPI, calcium-dependent and oxidizes both NADH and NADPH. This is a similar result to that obtained for the external-facing potato enzymes (Roberts *et al.*, 1995), which identified one calcium-independent, NADH specific enzyme, and one calcium-requiring enzyme exhibiting a preference for NADPH.

Two potato putative NDH2 cDNAs (identified by cDNA library screening using an *Arabidopsis* clone exhibiting homology to *S. cerevisiae* NDI1) were cloned (Rasmusson

[1] Two NDH2 genes according to this research (see 3.2.2)

et al., 1999). The two genes were sequenced, revealing 30-40% homology to NDI1, one of which included an EF-hand domain. This was provisionally termed 'NDB' (predicted mass 65kDa) while the other was termed 'NDA' (55kDa). Import experiments using mitochondrial preparations in conjunction with cross-linking assays showed that both expressed enzymes were imported, and that NDA was an external-facing and NDB an internal-facing enzyme. Surprisingly there was no evidence of target signal cleavage (unlike that observed with *S. cerevisiae*). These clones were also expressed in *E .coli* (see 1.5).

Antibodies raised against unconserved peptide segments of these expressed NDA and NDB clones (Rasmusson & Agius, 2001) were used in Western Blot analyses of both potato organelle fractions and mitochondrial preparations from various other plant species. The former identified a mitochondrial location for both enzymes (at the expected masses) but additional (weak) signals were detected for NDB in both peroxisome and chloroplast fractions. These were probably experimental artefacts as the positions corresponded to those for abundantly expressed proteins (a catalase protein, and the large Rubisco subunit respectively). Signals were detected in a number of other plant species (at the expected masses) though not for maize. A particularly strong signal was present in preparations from *Arum maculatum* spadices, a recognised source of highly expressed NDH2 protein (Cottingham & Moore, 1984). In 2D native/SDS-PAGE analysis of potato mitochondrial fractions incubated with the antibodies, bands were visible at higher masses which was interpreted as suggesting tri- or tetrameric association in the native state, but this observation has not been reported elsewhere.

The antibodies were also used in the examination of potato NDH2 expression levels in response to stress (Svensson *et al.*, 2002) and artificially controlled light levels (Svensson & Rasmusson, 2001). NDA expression in leaf mitochondria exposed for six days to a temperature of 5°C showed marked reduction in transcription (assessed by real time RT-PCR), protein abundance (immunoblotting) and NADPH oxidation rate. Transcription and protein expression of NDA was greatest in plant leaves in the early morning and lowest at night. Artificial subjection to extended dark conditions caused a decrease in expression, while a rapid rise was detected when restored to bright light.

In a more ambitious experiment the full-length potato NDB gene (now renamed 'St-**ndb1**') was transgenically expressed in tobacco using the well-established protocol of a CaMV 35S promoter and *Agrobacterium* transfection (Michalecka *et al.*, 2004). Second generation plants were assessed for transcription levels, protein abundance and NAD(P)H dehydrogenase activity broadly as described for the stress examination of endogenous NDB (above). Male sterility was observed in some progeny (probably a consequence of the unpredictability of the transfection insertion locus) and in some lines NADPH dehydrogenase activity was reduced ten-fold (compared to wild type), which was attributed to co-suppression of the endogenous NDH2. In other lines however proportionate increases of up to 2.3-fold were detected for mRNA, protein and NAD(P)H dehydrogenation activity. The activity was shown to be calcium-dependent, NADPH-specific and located to the external-facing side of the mitochondrial inner membrane.

Deposition of gene sequence data in publicly accessible databases has transformed the process of identifying putative proteins. The completion of genome sequencing projects has extended this approach to include the identification of complete enzyme families within a species. It has thus been possible to determine that there is a single NDH2 gene in *E. coli* and three in *S. cerevisiae* (though, of course, sequence homology alone is no guarantee of function). In this respect the completion of the *Arabidopsis* genome sequencing project (The Arabidopsis Genome Initiative, 2000) represented a landmark in plant research. The 25,498 genes in the publicly available sequence data (92% of the full complement of 115.4 megabases) is believed to include all protein-coding genes. Subsequently the rice genome has been added to the database (37,544 genes, 389 Mb, 95% complete, International Rice Genome Sequencing Project, 2005) and that for poplar (Tuskan *et al.*, 2006)[1].

Homology search of the *Arabidopsis* genome using (potato) St-**ndb1** as a query sequence has identified seven putative NDH2 genes (Michalecka *et al.*, 2003). These are classified into three groups: **nda** (2 genes), **ndb** (4 genes) and **ndc** (1 gene) as determined by sequence homology. The **ndb** type, previously identified in potato incorporates the putative calcium-binding EF-hand domain towards the C-terminus,

[1] Among landplants, the NCBI webiste lists Arabidopsis and rice genomes as 'complete', with those for poplar and grape at the 'assembly' stage. Several other species are listed as 'in progress' (April 2007) (http://www.ncbi.nlm.nih.gov/genomes/leuks.cgi).

while otherwise exhibiting close homology to **nda** genes, which in turn are 78% identical to potato **nda**s. **ndc**, also lacking the EF-hand domain, shows only 25% identity to **nda** genes and was more closely related (by sequence homology) to NDH2s of the cyanobacteria. A preliminary analysis of suc-gradient separated cell tissue using potato **nda** and **ndb** antibodies identified proteins of 48kDa and 61kDa respectively, consistent with the molecular masses of those for potato. Fusion constructs were then created combining the leader (targeting) sequence from *Arabidopsis* genes with GFP and these transformed into tobacco protoplasts. MitoTracker Red[1] was added to the preparation and examined using fluorescent microscopy. This showed that NDA1, NDA2, NDB1, NDB2 and NDC were located in the mitochondria (and not chloroplasts). No report is provided for NDB3 or NDB4 localisation.
In the same paper examination of transcript expression (real time RT-PCR) in response to changing light conditions showed that expression of **nda1** (but not **nda2**) is highly light-dependent though not in an expected manner. It is upregulated ten-fold 10 hours after being subjected to darkness, then decreases to a basal level (40% of the initial) after a further 24 hours in darkness. Within four hours of exposure to light it is again upregulated 5-fold. In contrast, the expression of NDA2 and NDB1 was not significantly affected by light conditions.

In a more detailed examination of light-dependent expression all seven putative NDH2s were examined along with the AOX and UCP genes (Escobar *et al.*, 2004). This showed that **nda1** and **ndc,** alone among the NDH2 genes, are upregulated in the presence of both short and long-term exposure to light, while **aox2** is downregulated. The two NDH2 genes were further examined by creating fusion products comprising the promoter/target regions of these with GUS[2], and transfecting *Arabidopsis* using the *Agrobacterium* vector. **nda1** was shown both to be localised, and to be light induced, in a pattern similar to that for photosynthetic genes (predominantly leaf mesophyll cells, expression as for Rubisco and Rubisco activase) and was thus proposed to be functionally associated with photosynthetic metabolism. The effect of light on **ndc** gene expression was less pronounced but was specific for red and/or blue wavelengths.

[1] A fluorescent marker for mitochondria
[2] This allows quantitative assessment by fluorometric analysis.

By examining an *Arabidopsis* **ndi1**$^-$ mutant strain, and by *in vitro* import experiments with isolated mitochondria it has also been determined that this gene encodes a matrix-facing NDH2 though it was not possible to determine whether it is specific for NADH or NADPH (Moore *et al.*, 2003).

NDH2s have been identified in other plant species though these have not been subjected to examination in comparable detail. A 58kDa external-facing NDH2 has been identified in beetroot (*Beta vulgaris*) (Cook-Johnson *et al.*, 1999) and that expression of this (or a similar) NDH2 is stimulated by slicing (Menz & Day, 1996a). An internal beetroot NADH dehydrogenase has also been proposed (Menz & Day, 1996b) though at just 43kDa this is somewhat less than the bacterial proteins (~47kDa); less likely still, on the basis of molecular mass, is that the mitochondrial NADH dehydrogenase isolated from maize at 32kDa (Knudten *et al.*, 1994) is in fact a member of this NDH2 family. However, NADH-Q_0 oxidoreductase activity (not NADPH) is reported for this purified enzyme and a similar activity has been shown for a 31kDa enzyme from beetroot (Luethy *et al.*, 1991), and 33kDa enzymes from *Arum maculatum* (Chauveau & Lance, 1991) and *Trypanosoma brucei* (Fang & Beattie, 2002a, and see below). Unfortunately no sequence data have been determined for these enzymes (and no recent publications have pursued this avenue) but the intriguing possibility exists that a further family of mitochondrial alternative dehydrogenases - 32kDa dimers, FMN-containing? - may yet be characterised.

1.5.5 NDH2 in other species

Biomedical research holds a particular interest for pathogen enzymes which are absent from the human proteome. Nowhere is this more manifest than with endemic diseases such as those caused by members of the *Plasmodium* and *Trypanosoma* species.
An enzyme was identified in *Trypanosoma brucei* located on the matrix-facing mitochondrial membrane which exhibited rotenone-insensitive NADH-quinone oxidoreductase activity (Fang & Beattie, 2002a). Surprisingly, in light of the contemporaneous NDH2 scholarship, this was determined by to be 33kDa protein (by

SDS-PAGE) containing non-covalently bound FMN, and predicted to be a homodimer *in vivo* on the evidence of filter chromatographic purification. (FMN was concluded on the combined evidence of fluorescence detection (excitation at 450nm→ detection at 525nm) and HPLC retention). It seems unlikely that a distinct mitochondrial NADH dehydrogenase has evolved in this species, though it is also unlikely that the purified sample was subject to a complete and exact bisection before SDS analysis *and* that the cofactor was misidentified. This group also focussed on the enzyme's apparent characteristic of elevated superoxide formation (Fang & Beattie, 2002b) based on a mechanism of a one-electron transfer producing semiquinone radicals - again not reported in other NDH2s which are assumed to transfer electrons (predominantly) in pairs. The group reported an ongoing task of cloning the responsible gene, but no further publication on this has been forthcoming at the time of writing[1]. This may, however, be a further example of a novel FMN family of NADH dehydrogenases (see above).

However this group did subsequently clone and express a 56.5kDa FAD-containing protein exhibiting characteristic NDH2 properties (Fang & Beattie, 2003).

[1] Subsequent publications from the Beattie group have addressed the alternative oxidase (aox), bc1 complex and human superoxide dismutase.

1.6 Elevated and heterologous NDH2 expression in *E. coli*

Attempts have been made to overexpress cloned genes from various species in *E. coli* but most do not report enzymatic characterisation of the purified protein. For example, *N. crassa* NDE1 was expressed with a His-tag in pQE-31 vector and the purified protein use for polyclonal antibody procurement (Melo *et al.*, 1999).

NDH2 has also served as a tool for examination of Complex I. *Paracoccus denitrificans* does not encode an endogenous NDH2, nor does it have a fermentation pathway so Complex I is an essential respiratory enzyme in this species. However Complex I was successfully inactivated by disrupting the *nqo8* and *nqo9* genes using the *ndh* gene (NDH2) from *E. coli* (Finel, 1996). This demonstrates transgenic rescue of ablated Complex I by NDH2. In a later publication (Björklöf *et al.*, 2000) it was also mentioned that the activity of the purified protein could be enhanced (though still with only modest activity) by the addition of FAD, suggesting that the prosthetic group could be incorporated into a fully synthesised and membrane bound peptide.

In an experiment to purify and characterise an overexpressed NDH2 (Björklöf *et al.*, 2000) the *E. coli ndh* gene was expressed in an *E. coli* host system. In spite of a high yield and absence of additional cofactor (FAD) the expressed protein attached to the periplasmic membrane rather than aggregating as inclusion bodies. Purification was achieved by chromatography (IMAC) and kinetic activity assessed with a variety of substrates demonstrating an apparent K_m of 34μM and V_{max} of 106 mmol NADH/min/mg with Quinone (Q_1) as acceptor. The purified protein however proved unstable requiring immediate concentration after chromatography. It was also sensitive to the type and concentration of detergent used to solubilise the membrane fraction (optimum: DM), and storage was possible only in the presence of glycerol and high ionic concentration at -80°C. The turnover rate (as high as 190 mmol/min/mg with Q_2 as acceptor) represents a 400-fold improvement over that obtained for the *P. denitrificans* experiments, an achievement that was attributed to refinement of the purification protocol.

Saccharomyces cerevisiae scNDI1 (the matrix-facing NDH2) was overexpressed in *E. coli* (Kitajima-Ihara & Yagi, 1998). The expression of three distinct fusion constructs

was attempted using the Novagen expression system: the full-length protein in pET-11a (T7-tag fusion), the full-length protein in pET-16b (His_{10}-tag fusion), and the mature protein in pET-24a (T7-tag fusion). Only the mature protein construct expressed successfully. This was purified by Triton X-100 solubilisation of the membrane fraction from ruptured cells followed by binding to a T7-affinity column. A protein of 57kDa was identified by SDS-PAGE and verified as NDI1 by peptide sequencing. Enzyme activity was assessed using the membrane fraction rather than purified protein and found to be between 3.5 and 12.5 times as active as that of a control (p-ET24a without cloned insert). The activity was inhibited by the addition of flavone, a recognised NDI1 inhibitor (de Vries & Grivell, 1988). Although not stated explicitly it may be inferred that the purified protein did not exhibit the expected activity.

The full-length NDH2 from *Trypanosoma brucei* was overexpressed in *E. coli* (Fang & Beattie, 2003) using the His-tag N-fusion construct in the pQE-30 plasmid (Qiagen). This produced a protein of 56.5kDa (SDS-PAGE) that, by inference[1], was expressed (surprisingly) as a soluble protein which was then purified on a nickel column. Enzyme assays were performed using NADH and three quinone acceptors (Q_0, Q_1 and duroquinone) achieving a turnover of 1.8 μmol/mg/min using Q_1. The rates for deamino-NADH and NADPH oxidation were considerably lower (57 and 64 nmol/mg/min respectively).

Potato NDA (matrix-facing) and NDB (calcium-dependent, external-facing) proteins were overexpressed in *E. coli* as full-length proteins with an N-terminal fusion Trx- and his-tags within the expression vector pET-32 (Novagen; see Chapter 3 for details of this construct)(Rasmusson *et al.*, 1999). SDS-PAGE analysis indicated proteins of 73kDa and 84 kDa for NDA and NDB fusion products respectively (roughly as predicted). NDA was partially located in the bacterial membrane (released by Triton-X100 solubilisation) and partly as inclusion bodies (denatured in urea), while the NDB fusion product was purified as a soluble protein by a high concentration of monovalent salt or EDTA alone, suggesting only weak (and/or no) membrane association in spite of the close homology to NDA at the C-terminal domain (assumed to be involved in membrane binding; see also Chapter 4). There is no report of enzyme kinetic characterisation of the expressed protein in this paper.

[1] the paper is not explicit but makes no mention of detergent solubilisation.

1.7 *In silico* protein structure modelling

In addition to the homology comparisons of NDH2 sequence data and phylogenic analyses contained in the above publications, a putative structure for the *E. coli* NDH2 has been created using the 'homology modelling' technique (Schmid & Gerloff, 2004). The template used for this was the NADH peroxidase from *Streptococcus faecalis* (PDB structure: 2NPX) to which it bears the closest sequence homology among available resolved structures. A significant difference between these, and indeed other members of the "Class II pyridine nucleotide-disulfide oxidoreductase family"[1], is that they are predominantly soluble homodimers *in vivo* and the C-terminal region (typ. 20%) serves as a dimerisation domain. NDH2s are of roughly the same overall length as these but differ in that they are membrane-bound monomers so, not surprisingly, the C-terminal regions of these bear scant primary or secondary structural homology to members of this family. It is proposed in this paper that NDH2s should be classified as a new subgroup of this family and that the C-terminal region plays a role in binding to the membrane. Consequently this domain has been omitted form the model (PDB structure: 1OZK); analysis of its secondary structure has been used to propose membrane interactions, though not (consistent with other reports) the presence of any putative transmembrane helices.

[1] pfam classification of protein domains

1.8 Therapeutic applications

A plant or fungal enzyme that has no mammalian paralogue is not an obvious target for biomedical therapeutic research. It comes then as something of a surprise to discover that NDH2 is being actively pursued as a candidate for gene therapy. Mitochondrial genetic defects, whether inherited, acquired though exposure to toxic environmental conditions or the natural effect of ageing, are associated with various diseases. Those specific to the respiratory chain components can cause encephalomyopathies[1] (DiMauro & Schon, 2003). The neurodegenerative disorder, Parkinson's disease, has been found to be associated with defects in Complex I, the symptoms (as well as post-mortem disease distribution patterns) of which are similar to those of animals exposed to rotenone (Sherer *et al.*, 2003). A causal link between genetically compromised Complex I subunit(s) and Parkinson's remains unproven but is sufficiently compelling to warrant continued research.

The delivery of genetic material to tissue employing disabled adenovirus vectors carrying donor DNA, "gene therapy", is an established yet evolving technology (Smith *et al.*, 2003; Zolotukhin *et al.*, 2002). The *Chlamydomonas reinhardtii* mitochondrial ATPase subunit 6 gene has been demonstrated to recover a human cell line cell containing a pathogenic mutation in the ATP6 gene (Ojaima *et al.*, 2002). Remarkably the algal mitochondrial targeting signal was correctly recognised and cleaved after import though the observed growth rates were ten-fold reduced when compared to that of wild type cells. Complex I itself is an inappropriate target for such therapy due to its multiplicity of subunits and issues surrounding stoichiometric expression. The single peptide NDH2, however, performs the essential function of Complex I (albeit without optimum energy conservation) and its gene is a suitable size for this delivery system (Yagi *et al.*, 2001). Experiments to date have employed the *S. cerevisiae* NDI1 gene which was first shown to be imported correctly by cotransfection to the matrix-facing side of the inner membrane of Complex I deficient Chinese hamster cells (Boo Seo *et al.*, 1998). It has subsequently been shown to confer viability on human kidney cells in the presence of rotenone (Boo Seo *et al.*, 1999) and in mammalian dopaminergic nerve cell lines (Boo Seo *et al.*, 2002).

[1] degenerative effects in brain and muscle tissue.

1.9 Aims of the current research

The purpose of the current work was to broaden the understanding of NDH2 structure and function using a combination of computational, analytic and biochemical approaches. The strands of the research, presented in this thesis, may be summarised as follows:

A manual method was used to align sequence data from available putative and confirmed NDH2 data. Homology at conserved loci (where possible, loci unique to NDH2s) was employed in order to group the enzymes into subfamilies and identify regions and motifs common to NDH2s and to the individual groups (Chapter 3).

Using the results of this analysis a putative structure of the FAD-containing *E. coli* NDH2 was created using homology modelling, and from this a model of the *Arabidopsis* **nda1** enzyme was created. This process included examination of related enzymes in the structural database (PDB) in order to distinguish conserved and enzyme-specific loci. From inspection of the model, particularly the 3D positions of residues critical for substrate binding, the relevance of specific conserved residues and motifs in the primary structure were determined. By extension, this affords prediction of functional properties of NDH2s in general from their sequence data alone.
An understanding of structural features from the model provides information beneficial to the biochemical isolation of active protein (Chapter 5).

cDNA clones of putative NDH2 genes were created from *Arabidopsis* leaf tissue using the techniques of Molecular Biology. These were subcloned into expression vectors with terminal His-tags for heterologous expression in *E. coli*, in order to purify and characterise active enzymes. Protein preparation obtained in this manner may be suitable for crystal growth, from which X-ray crystallographic analysis can provide a 3D structure of the enzyme finally resolving the question of membrane interaction and, hopefully, confirming the validity of the computationally derived model (Chapter 4).

2 Materials and Methods

2.1 General information

Except where otherwise identified, enzymes were obtained from New England Biolabs (NEB) (Hertfordshire, UK) and all other reagents from Sigma-Aldrich (Dorset, UK). Custom oligonucleotides were obtained from MWG-Biotech (Germany). Double distilled water (ddH_20) was provided by a Maxima water filtration unit (ELGA, Buckinghamshire, UK). All nucleotide preparations were stored at -20°C.

Where appropriate, glassware, solutions and media were sterilised either by autoclaving (15 psi,. 121°C, 15 min) or by filter sterilisation using a syringe and 0.22 µm filter (Millipore, Hertfordshire, UK). Aseptic techniques were used throughout. Bacterial cultures were destroyed by the addition of bleach (~20%, 30 mins), and agar plates by autoclaving prior to incineration, in accordance with approved practices for the disposal of genetically modified organisms.

2.1.1 Stock reagents

0.5M EDTA pH 8.0

372g Na_2EDTA was dissolved in 800ml ddH_20 by stirring over hot plate. The pH was adjusted to pH 8.0 by the gradual addition of 10M NaOH and made up to 1.0l with further ddH_20. The solution was autoclaved and stored at RT.

TAE

A 1l stock of X50 TAE (240g Tris-HCL, 10ml 0.5M EDTA, 57.1ml glacial acetic acid, made up to 1.0l with ddH_20) was stored at RT. From this, 50-fold dilutions in ddH_20 were made to provide X1 TAE, also stored at RT.

Ampicillin 100mg/ml

1g ampicillin powder was dissolved in 10ml ddH_20, filter sterilised, and stored in 500µl aliquots at -20°C for up to three months.

IPTG 1M

3.125g was dissolved in 10ml ddH_20 and stored at –20°C for up to three months.

X-gal 20mg/ml

200µg X-gal was dissolved in 10ml dimethylformamide and stored at -20°C. The tube was wrapped in aluminium foil to minimize exposure to light.

50% Glycerol

Glycerol was dissolved in ddH_20 to 50% (v/v), autoclaved and stored at RT.

Luria-Bertani (LB) medium

1% (w/v) Bacto-tryptone (Difco, MI,USA) 0.5% (w/v) bacto-yeast extract (Difco, MI, USA) 0.17M NaCl.. The medium was autoclaved and stored at RT (Sambrook *et al.*, 1987).

Luria-Bertani (LB) solid medium

LB medium was supplemented with 1.5% (w/v) bacto-agar (Difco, MI,USA), autoclaved andstored at RT.

10mM NADH

53mg $NaCO_3$ was dissolved in 10ml ddH_20 to provide a stock 50mM $NaCO_3$ solution stored at RT. 7.09mg NADH was dissolved in 1ml 50mM $NaCO_3$ and stored at –20°C for up to one month.

1mM FAD

8.3 mg was dissolved in 10ml ddH_20, filter sterilized and stored in 1ml aliquots at –20°C.

100mM FeCN (Potassium Ferricyanide)

329mg KFeCN was dissolved in 10ml ddH_20 and stored at –20°C wrapped in aluminium foil to minimise exposure to light.

6M Guanidine

57.1g guanidine was dissolved in 100ml H_20 and stored at RT.

8M Urea

48.0g urea was dissolved in 100ml H_20 and stored at RT.

DEPC treated Water

1ml DEPC was added to 1.0l ddH_20, left to stand for an hour then autoclaved. DEPC water was stored at RT.

2.2 Nucleic Acid procedures

2.2.1 Isolation of RNA from *Arabidopsis* leaves

Arabidopsis RNA preparations were performed using the Qiagen Plant RNeasy system with 100g leaves per reaction. In view of the potentially low concentration of NDH2 mRNA species, and the relative length of mature transcripts (up to 2000 nucleotides), particular attention was paid to the preparation of equipment to minimise the effects RNA degradation. (RNAse enzymes are typically *not* inactivated by autoclaving, ethanol treatment or other conventional sterilisation methods.) All apparatus was washed with detergent rinsed with ddH_20 then washed with ethanol and rinsed again. It was then washed with RNaseAway (E&K Scientific #EK-3352, CA, USA), a detergent including a cocktail of additional reagents with known RNAse inhibitory properties, and rinsed with DEPC water. Pipettes were dismantled prior to washing; separately supplied RNase-free tubes and pipette tips were used throughout, and all prepared apparatus was wrapped in aluminium foil and stored away from other equipment. Work surfaces, vortex and microcentrifuge machines were also washed with RnaseAway and rinsed with DEPC water.

A ceramic mortar and pestle were similarly washed, wrapped in aluminium foil, autoclaved and stored overnight at –80°C as tissue disruption is performed in liquid nitrogen. Latex gloves were changed at five minute intervals (as suggested in the manufacturer's instructions) and whenever in accidental contact with non-prepared surfaces. All reagents were maintained on ice throughout the procedure and all steps performed promptly.

Liquid nitrogen was placed in the mortar and the latter allowed to cool further until the boiling was minimal. 100g *Arabidopsis* leaf material (stored at –80°C) was weighed, transferred to the mortar and ground vigorously for five minutes, liquid nitrogen being added to balance the evaporation. Residual liquid nitrogen was the allowed to evaporate and the disrupted tissue transferred to a 2ml tube, 450µl 'RLT' lysis buffer (containing GITC) added and the tube vortexed to release the cytoplasm. The mixture was transferred to a filter column and centrifuged for two minutes. Clear filtrate was then transferred to a fresh tube, 225µl 99% ethanol added and mixed by pipetting, and the sample loaded on to a silica-gel membrane column that binds RNA (but not DNA). This was centrifuged for 15 seconds, 700µl buffer RW1 added and centrifuged again to remove DNA. The column was then washed twice by the addition of buffer 'RPE' (80% ethanol) followed by centrifugation (15 seconds, then 2 minutes). After a final 60 second centrifugation in a fresh collection tube to remove residual ethanol, RNA was eluted with 50µl RNase-free water.

2.2.2 Isolation of mRNA from total RNA preparations

mRNA was recovered from total RNA preparations using the Qiagen Oligotex kit (W. Sussex, UK) following the manufacturer's instructions. This uses the $dC_{10}T_{30}$ oligonucleotides covalently linked to polystyrene particles which bind poly-A tails of eukaryotic mRNA under specified conditions. ~50µg total RNA was loaded to the matrix yielding ~2µg mRNA in a 50µl elution, assuming 95% efficiency and 4% mRNA in total RNA.

.

2.2.3 Reverse Transcription PCR

Reverse transcription was performed using a combination of Qiagen Omniscript and Sensiscript systems with the addition of RNaseOUT (Invitrogen #10777-019, 40U/µl) to minimise RNA degradation.

Template and all reagents were thawed on ice and a 5µl aliquot of the RNase inhibitor diluted in 15µl RNAse-free H_20 (final 10U/µl). 20µl reactions were made by mixing 2µl X10 RT buffer, 0.5mM each dATP, dCTP , dGTP, dTTP, 0.5U/µl RNsaOUT, 1µM primer (see text), 0.2U/µl RT enzyme (see text), 2µg template, and RNase-free H_20 to 20µl. The template was added after the remaining reagents had been briefly mixed and centrifuged and the reaction incubated at 37°C for 60 minutes. The reaction mixture was used immediately and/or stored at –20°C. PCR was performed using either Taq or VENT polymerases (see text).

2.2.4 Isolation of plasmids from *E. coli*

Plasmid purification was based on the Qiagen Miniprep alkaline lysis system (W. Sussex, UK) yielding ~10µg DNA per reaction. All centrifugation was performed on a Biofuge Pico bench-top microcentrifuge (Heraeus Instruments; Osterode, Germany), For the high-copy cloning plasmids (XL1, Bluescript, pGem) 2ml culture (typically overnight) was harvested by centrifugation for five minutes and resuspended in 250µl buffer (50 mM Tris-HCl, pH 8.0; 10 mM Na_2-EDTA; 100 µg/ml RNase A) to which 250µl lysis buffer (200 mM NaOH; 1% (w/v) SDS) was added, and lysis was allowed to proceed for three minutes at RT. 350µl neutralising buffer (3 M KOAc, pH 5.5) was added, the mixture incubated on ice for ten minutes, and then centrifuged for a further ten. The supernatant was loaded on to an ion exchange column and centrifuged for one minute. For lower-copy plasmids (pET vectors) the above procedure was performed in triplicate and all three lysate supernatant phases loaded to a single column. The column was then washed with 750 µl buffer 'PE' (95% ethanol) – once for high- and twice for lower-copy plasmids – and subjected to a final one minute centrifugation to remove trace ethanol. Purified plasmid was generally eluted in 40µl 10mM Tris-Cl pH 8.5), but in 50µl where gel extraction was subsequently to be performed.

2.2.5 DNA extraction from agarose gel

DNA fragments were excised from 0.8% agarose gels viewed on a UV trans-illuminator (Ultra-Violet products, CA, USA) using a razor blade. DNA was recovered from the agarose by the agarase digestion method using the Qiagen (W.Sussex, UK) QuiQuick Get Extraction Kit in accordance with the manufacturer's instructions. Typically 30% recovery was achieved, as estimated by gel visualisation.

2.2.6 DNA restriction endonuclease digestion

Typically 2μg DNA was subjected to digestion in a 20μl reaction volume by 0.5μl enzyme (10U typ.) with the manufacturer's recommended buffer and BSA as required, for one hour at 37°C. Following digestion the enzyme was inactivated by heat treatment where possible, or by ion exchange chromatography using the Qiagen (W.Sussex, UK) Nucleotide Removal Kit. Where appropriate (ie. other than PCR terminal digests) a sample of the digest was examined on a 1% (w/v) agarose gel to confirm complete successful endonuclease activity.

2.2.7 DNA purification

Following restriction enzyme digests and other procedures, DNA was purified form nuclease enzymes, terminal nucleotides and buffer using the QIAquick Nucleotide Removal kit (Qiagen, W.Sussex, UK) following the manufacturer's instructions.

2.2.8 DNA terminal phosphatase

In order to minimise unwanted DNA ligation products (typically re-circularisation of digested vector plasmid), where appropriate the 5-'terminal phosphate group was first removed using Antarctic Phosphatase (NEB #M0289S). Typically 5U enzyme (1µl) was incubated with 10ng vector in a 20µl reaction containing the optimum buffer (50mM Bis-Tris-Propane-HCl, 1mM $MgCl_2$, 0.1mM $ZnCl_2$, pH 6.0) for 10 minutes at 37°C, though other buffers were also used (see text).

2.2.9 DNA ligation

The Quick Ligation system (NEB #M2200S) was used for all DNA ligation operations following the manufacturer's recommended instructions, except that while a 3:1 (insert:vector) molar ratio of 'ends' was used with unphosphatased vectors, a 1:1 ratio was used where the vector had had the 5'-terminal phosphate groups removed (see above). The enzyme in this system is the conventional T4 ligase but the inclusion of PEG-6000 in the reaction buffer significantly enhances reaction time. Typically 100ng DNA was incubated in 20µl reaction buffer (132mM Tris-HCl, 20mM $MgCl_2$, 2mM dithiothreitol, 2mM ATP15% PEG 6000, pH 7.6) including 1µl ligase for 10 minutes at RT.

2.2.10 Polymerase Chain Reaction (PCR)

DNA was amplified from Reverse Transcription product or plasmid template by the Polymerase Chain Reaction using a Techne Genius thermocycler (Cambridge, UK). Both VENT polymerase (a proofreading enzyme) and Taq were used (see text) in 50µl reactions containing Thermopol buffer #B9004S (20mM Tris-HCl pH 8.8, 10mM KCl, 10mM $(NH_4)SO_4$, 2mM $MgSO_4$, 0.1% Triton X-100), 800µM dNTP (200µM each dATP, dCTP, dGTP and dTTP), 1µM of each primer, 10ng template DNA and 0.5U (VENT) or 2U (Taq). With VENT amplification additional parallel PCR reactions were performed with 4mM and 6mM $MgSO_4$. The reaction mixture (excluding polymerase) was vortexed to mix and the polymerase was added only after the initial cycle to 95°C.

The following default thermocycle schedule was used:

1)	95°C	5 minutes	
2a)	95°C	1 minutes	} 25 cycles
2b)	note 1	1 minutes	
2c)	72°C	note 2	
3)	72°C	4 minutes	
4)	4°C	indefinite	

[1] annealing temperature depends on primer
[2] depends on product length and polymerase (VENT 0.5k/min, Taq 2k/min)

Following an initial inspection on agarose gel selected PCR products were purified using the Qiagen PCR CleanUp kit using the manufacturer's recommended instructions prior to further processing.

2.3 Biochemical and general procedures

2.3.1 Centrifugation

Except where otherwise stated all benchtop microcentrifugation was performed at 13000 rpm (16000g) in a Heraeus Instruments Biofuge Pico microcentrifuge (Osterode, Germany)

2.3.2 Growth of bacterial colonies

Luria-Bertani (LB) solid medium was heated to 70°C in a water bath, antibiotics (and/or other supplements) added as required, the medium poured into plastic petri plates and allowed to solidify.

Plate types:	LB	Lb medium only (stored at RT)
	LB-Amp	LB with 50 µg/ml ampicillin (stored at 4°C).
	LB-blue	LB with 50 µg/ml ampicillin, 50 µg/ml X-gal and 10 µg/ml IPTG (stored at 4°C).

2.3.3 Bacteria cultures

Cells were picked from plates using a sterilised pipette tip, or removed by pipette from culture, and used to inoculate fresh LB medium. All bacterial cultures were grown at 37°C in a shaker. Where appropriate ampicillin was added to 50μg/ml.
"Minicultures" were grown in 3ml LB medium in sealed 10ml plastic Sterilin bottles (Staffordshire, UK). Larger scale cultures were grown in glass conical vessels with loose cotton wool stoppers. 50ml cultures were grown in 200ml flasks, 500ml in 2l flasks.

2.3.4 Glycerol stocks of bacteria

For long-term storage of bacterial strains culture was mixed with an equal quantity of 50% glycerol (25% (v/v) final), vortexed to mix and snap-frozen in liquid N_2. Aliquots (500ml) were stored at -80°C.

2.3.5 Preparation of Competent *E. coli* Cells

E. coli cuture (strain DH5α) was prepared for transformation by the $CaCl_2$ method adapted from Sambrook, 1989. 10ml LB in a 20ml Sterilin tube was inoculated with 100μl freshly cultured stationary phase cells and incubated for 3 hours at 37°C (~10^8 cells/ml). Cells were harvested in an Eppendorf 5403 centrifuge (Hamburg, Germany) at 5000g for 5 minutes at 4, the medium removed, the cell pellet resuspended in 10ml chilled 100mM $MgCl_2$ and left on ice for ten minutes. The cells were harvested again as described above and resuspended in 1ml chilled 100mM $CaCl_2$ by gentle swirling and incubated on ice for 30 minutes. Competent cells were left for 24hrs at 4°C prior to use in transformation.

2.3.6 Plasmid Transformation of Competent *E. coli* Cells

Competent DH5α cells were transformed by the 'heat shock' method adapted from Cohen (1972). 100μg DNA (typically 1μl plasmid) was placed in a chilled Eppendorf tube with 100μl competent cells and incubated on ice for 30 minutes. The tube was then transferred to a 42°C water bath for 90 seconds and returned to the ice for a further 5 minutes. 500μl LB was then added to the tube, mixed, and incubated at 37°C for one hour. From this a 40μl aliquot was pipetted on to LB-agar plates containing 50μg/ml ampicillin and spread evenly with an ethanol-sterilised glass rod. The remaining cells were then harvested in a microcentrifuge for two minutes (16000g), resuspended in 40μl LB by pipetting and transferred to a second LB-agar plate. The two plates (low- and high-concentration) were left to incubate at 37°C for around 16 hours prior to visual inspection for the growth of colonies. Plates were stored for up to a month at 4°C where appropriate and/or colonies selected for culture as described above.

2.3.7 Growth of *Arabidopsis* plants

Seeds were sterilised by soaking in bleach for ten minutes (6% NaOCl, 0.05% Tween-20) and then washed five times in ddH_20. Individual seeds were then picked using a pasteur pipette and placed in a petri dish into the surface of MS-agar medium (0.5X MS (Murashige and Skoog salts: 20.5mM NH_4NO_3, 3mM $CaCl_2$, 1.5mM $MgSO_4$, 19mM KNO_3, 1.25% KH_2PO_4), 0.8% agar, autoclaved). The petri dish was sealed with parafilm and kept at 4°C for three days before being placed in a high light (~150 $\mu Em^{-2}s^{-2}$) environment at 24°C. When roots had developed to around 3cm individual plants were transferred to 8cm pots containing fertiliser and these placed on a water tray in a controlled environment (14hours light / 10 hours dark, 150 $Em^{-2}s^{-2}$, 24°C). Plants were harvested once they had grown to a height of around 10cm.

2.4 Analytic techniques

2.4.1 Agarose gel electrophoresis

All DNA and RNA gels were made using the Minigel Submarine HE-33 system (Hoefer Pharmacia Biotech, CA, USA) with a constant voltage provided by a PowerPac PAC300 (Bio-Rad, CA, USA). 50ml 1% (w/v) gels were made by dissolving 50mg agarose in 50 ml TAE, heated to boiling point in a microwave. (For subsequent DNA gel extraction applications, 0.8% gels were used.) On cooling to ~70° ethidium bromide was added (0.1µg/ml), swirled and poured into the mould. A twelve- or eight-well comb was inserted and the gel allowed to cool to RT. Samples were prepared by the 5:1 (v/v) addition of X6 Loading Dye (#G1881, Promega, CA, USA) and pipetted into the wells. A DNA marker solution was made mixing five parts ddH_20 with one part GeneRuler DNA markers (#SM0311, Fermentas) and one part X6 Loading Dye (#G1881, Promega, CA, USA). 6µl of this marker solution was loaded to additional lane(s) in the gel to assess sample DNA fragment size. At this concentration the 2000bp band contains 184ng DNA (6µl) allowing estimation of sample concentration. The tank was filled with TAE running buffer and the power supply set to 100V. Images were photographed under UV (DOC-IT UVP Gel Documentation System).

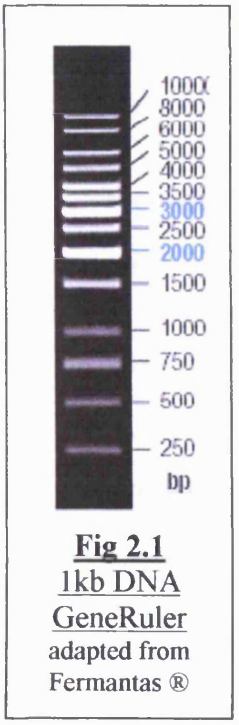


Fig 2.1
1kb DNA GeneRuler
adapted from Fermantas ®

2.4.2 SDS polyacrylamide gel electrophoresis (SDS-PAGE)

All protein electrophoresis was performed using the SDS-PAGE method (adapted from Laemmli, 1970) using the Mini-Electrophoresis apparatus (Bio-Rad, CA, USA) with a constant 100V provided by a PowerPac 300 (Bio-Rad, CA, USA). Protein bands were separated in 10% acrylamide gel (29:1 acrylamide/bis-acrylamide).

Components were thoroughly washed with detergent and then methanol prior to assembly. The 10% 'resolving' gel was prepared (10% acrylamide, 0.375M Tris-HCl pH 8.8, 0.1% (w/v) SDS, 0.05% (w/v) TEMED, 0.075% (w/v) APS) and poured between the plates leaving a 1cm gap at the top. The TEMED and APS were added

immediately prior to pouring. A layer of butane was placed above this to ensure a flat gel surface and allowed to set (about one hour). The butane was then removed and the 1cm gap thoroughly washed with dH_20 and dried using absorbent paper. A 3.3% 'stacking' gel (3.3% acrylamide, 0.125 M Tris-HCl ph 6.8, 0.1% (w/v) SDS, 0.05% (w/v) TEMED, 0.075% (w/v) APS) was then poured into the 1cm space and a ten-well comb inserted. The comb was removed once the second gel layer had set, the resulting wells washed in dH_20, and the plates assembled in the resolving tank equipment and reservoirs filled with 'running buffer' (250mM Tris-HCl pH 8.3, 0.1% (w/v) SDS, 1.2% (w/v) glycine) ensuring that no air bubbles remained in the wells.

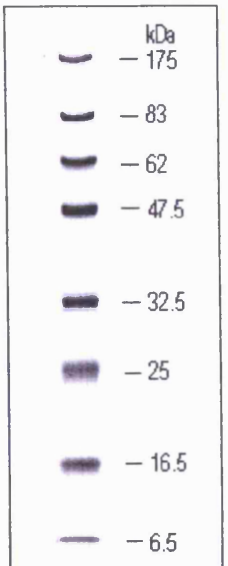


Fig 2.2
#P7708
Mol. Weight
Markers
adapted from
NEB ®

Samples (10ml) were prepared by mixing with an equal volume of 'sample buffer' (0.0625M Tris-HCl pH6.8, 2% (w/v) SDS, 20% (v/v) glycerol, 5% (v/v) β-mercaptoethanol, 0.1% (w/v) bromophenol blue) and boiled in a water tank for ten minutes. The 20µl prepared samples were then pipetted into the wells and additional lane(s) loaded with 5µl broad range prestained protein molecular markers, obtained from either NEB (#P7708) or Promega (#V8491). Electrophoresis was performed at a constant 100V and allowed to continue until the dye-front reached the bottom of the plates (or as required depending on the bands of interest).

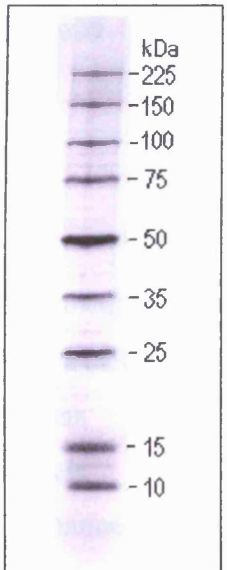


Fig 2.3
#V8491
Mol. Weight
Markers
adapted from
Promega ®

For visualisation of the protein bands the gels were removed from the plates and placed in plastic dishes containing 'stain buffer' (1% (w/v) Coomassie blue, 20% (v/v) methanol, 20% (v/v) acetic acid), heated to boiling in a microwave and left for two hours on a moving surface. The gels were then washed in dH_20 and placed in ~20ml 'destain buffer' (10% (v/v) methanol, 10% (v/v) acetic acid), heated to boiling in a microwave and placed on the moving table. After ten minutes this destain buffer was replaced with fresh destain buffer, heated and left on the moving table for typically four hours. The gel image was recorded by capturing on a flatbed scanner and saving on computer disc.

2.4.3 Cell disruption

The cytoplasm of *E. coli* cells was released by the French Press method, whereby cell walls are ruptured by the sudden release of pressure. A Thermo FRENCH® Pressure Cell Press was adjusted to a hydraulic pressure of 1280 psi, resulting in a pressure within the "40K" cell of 20,000psi. The pressure cell was placed on ice for ten minutes prior to filling with the cell sample (typically 40ml). The cell was placed in the press with the collection tube placed in ice. Hydraulic pressure was applied and the release valve adjusted for a flow rate of approximately one drop every five seconds. Each sample was subjected to two passes through the pressure cell.

2.4.4 Bradford Assay

Protein concentration was estimated using Bradford Reagent (Sigma #B-6919). BSA (NEB) was diluted to 1mg/ml as determined by an absorbance of 0.66 at 280nm (A_{280} = 0.6) and then further diluted to provide standards of 10, 20, 40, 60, 80 and 100µg/ml. Cuvettes samples were prepared by the addition of 100ml BSA standard or sample to 1ml Bradford reagent, mixed, and left to stand for five minutes. Absorbance was measured at 595nm and concentration determined from BSA standard curve.

2.4.5 Determination of NADH Dehydrogenase activity

Enzyme kinetics was determined by time-course spectrophotometric measurement of NADH concentration in a reaction using potassium ferricyanide (FeCN) as the electron acceptor. To improve NADH specificity a two-channel spectrophotometer was employed (Sigma ZWS11 Zweiwellenlängenspektralphotometer) with channels configured for 340nm (A channel: NADH peak absorption) and 380nm (B channel: background). Reactions were performed in quartz cuvettes using the tungsten light source. The difference output (A-B) was connected via a custom designed analog-to-digital converter to a computer where samples were recorded at 4Hz and displayed by a PC programme written for this task (fig 2.4, see also Appendix A.2).

A 1ml quartz cuvette was first thoroughly rinsed with ddH_20 then filled with 950µl KPi buffer (pH8.0). To this 10µl FeCN (100mM) was added (final: 1mM) and mixed by stirring with a plastic stick. The cuvette was placed in the spectrophotometer and the [A-B] output adjusted for a zero reading. 10ml NADH (10mM) was then added (final: 100µM), stirred, and the absorbance recorded (typically 0.3). As a positive control 1µl KH particles (Keilin-Hartree sub-mitochondrial preparation from bovine heart at 10mg/ml protein, stored at –80°C) was added prior to each assay exercise to verify reagent integrity and spectrophotometer operation. The above FeCN and NADH concentrations were first determined empirically such that these were not limiting factors in the assay.

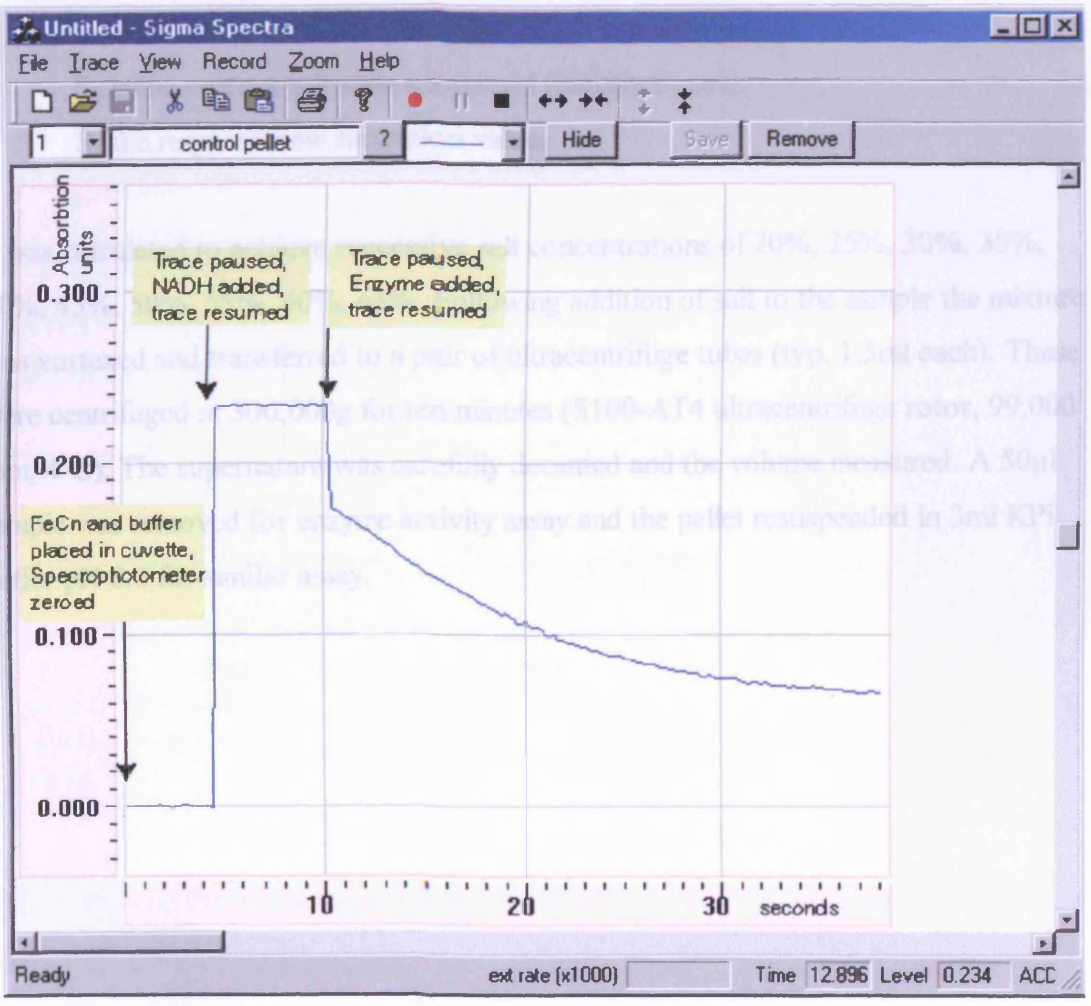


Fig 2.4 Time-course assay of NDH2 enzyme activity

Screen-shot of 'Spectra' (see Appendix A.2) recording of NADH concentration against time.
Vertical Axis: Differential Absorption Units (A_{340} - A_{380}). Horizontal axis: time in seconds.
Note: recording is paused for approximately five seconds during addition and mixing of NADH and enzyme samples (marked).

2.4.6 Salt Precipitation of detergent solubilised protein

Salt precititation experiments were performed by the serial addition of either Ammonium Sulphate or Ammonium Acetate to a solution of mitochondrial cell fragments in 0.5% (w/v) deoxycholate, 0.1M KPi pH8.0. Saturated salt solution was added to the sample according to the formula

$$\mathbf{v} = V_1 (S_2 - S_1)/(100 - S_2)$$

where **v** is the volume of saturated salt solution to add,

V_1 is the volume of the detergent sample (initially 3ml),

S_1 is the current salt saturation level (initially 0), and

S_2 the required new saturation value.

v was calculated to achieve successive salt concentrations of 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%. Following addition of salt to the sample the mixture was vortexed and transferred to a pair of ultracentrifuge tubes (typ. 1.5ml each). These were centrifuged at 300,000g for ten minutes (S100-AT4 ultracentrifuge rotor, 99,000 rpm, 4°C). The supernatant was carefully decanted and the volume measured. A 50µl sample was removed for enzyme activity assay and the pellet resuspended in 3ml KPi buffer pH 8.0 for similar assay.

2.5 *In silico* modelling methods

All homology modelling was performed using version 8.1 of Modeller[1] (Marti-Renen, 2000). The modelling approach is detailed in Chapter 5. The principal files used in this procedure are provided here and may also be found in the Supplementary Data CD.

The following sequence, ecCx.ali, was used to define the C-terminal deleted *E. coli* NDH2:

```
>P1;ecCx
sequence:ecCx:::::::0.00:0.00
LTTPLKKIVIVGGGAGGLEMATQLGHKLGRKKKAKITLVDRNHSHLWKPLLHEVATGSLDEGVDALSYLAHARNH
FQFQLGSVIDIDREAKTITIAELRDEKGELLVPERKIAYDTLVMALGSTSNDFNTPGVKENCIFLDNPHQARRFH
EMLNLFLKYSANLGANGKVNIAIVGGGATGVELSAELHNAVKQLHSYGYKGLTNEALNVTLVEAGERILPALPPR
SAAAHNELTKLGVRVLTQTMVTSADEGGLHTKDGEYIEADLMVWAAGIKAPDFLKDIGGLETNRINQLVVEPTLQ
TRDPDIYAIGDCASCPRPEGGFVPPRAQAAHQMATCAMNNILAQMNGKPLKNYQYKDHGSLVSLSNFSTVGSLMG
LTRGSM*
```

The initial attempts to model using a chimeric alignment used the `automodel` class with two templates 1FL2fitXH and 1HXC, the former being a reorientation of 1FL2 to superimpose it upon 1XHC:

```
from modeller.automodel import *

env = environ()
env.io.atom_files_directory = './:../../pdb_files/'

a = automodel(env, alnfile='ecCx-XH1.ali',
              knowns=('1FL2fitXH','1XHC'), sequence='ecCx')
a.starting_model = 1
a.ending_model = 5
a.make()
```

This referenced the following alignment file in which the overlap region (β_8) is shown in red:

[1], http://salilab.org

```
>P1;1FL2fitXH
structureX:1FL2fitXH:   212 :A: 307 :A:undefined:undefined:-1.00:-1.00
----AYDVLIVGSGPAGAAAAIYSARKG-----IRTGLMGERFGGQILDTVDIENYISVPKTEGQKLAGALKVHVD
EYDVDVIDSQSASKLIPAAVEGGLHQIET-----------------------------------
----------------------------------------------------------------
----------------------------------------------------------------------
----------------------------------------------------------------
----------------------------------------------------------------*

>P1;1XHC
structureX:1XHC:  83 :A: 351 :A:undefined:undefined:-1.00:-1.00
----------------------------------------------------------------------
------------------------KVVIT-----------EKGEVPYDTLVLATGARAR-EPQI-
KGKEYLLTLRTIFDADRIKESIENS-------------GEAIIIGGGFIGLELAGNLAEA-----
----------GYHVKLIHRGAMFLGL-DEELSNMIKDMLEETGVKFFLNSELLEANEEGVLTNS-GFIEGKVKIC
AIGIVPNVDLARRSGIHTGR--GILIDDNFRTS-AKDVYAIGDCAEYSGIIAGT---AKAAMEQARVL
ADIL-----KGEPRRYNFKFRSTVFKFGKLQIAIIGNTKGEGKWIEDNTKVFYIGAVVFNDIRKATKLE*

>P1;ecCx
sequence:ecCx:     : :     : ::: 0.00: 0.00
LTTPLKKIVIVGGGAGGLEMATQLGHKLGRKKKAKITLVDRNHSHL--WKPLLHEVATGSLDEGVDALSYLAHARN
HGF-QFQLG-SVIDI-DREAK---TITIAELRDEKGELLVPERKIAYDTLVMALGSTSN-DFNTP
GVKENCIFLDNPHQARRFHQEMLNLFLKYSANLGANGKVNIAIVGGGATGVELSAELHNAVKQLH
SYGYKGLTNEALNVTLVEAGERILPALPPRISAAAHNELTKLGVRVLTQTMVTSADEGGLHTKDGEYIEADLMVW
AAGIKAPDFLKDIGGLETNRINQLVVEPTLQTTRDPDIYAIGDCASCPRPEGGFVPPRAQAAHQMATC
AMNNILAQMNGKPLKNYQYKDHGSLVSLSNFSTVGSLMGNLTRGSM*
```

In the alignment file for the initial chimeric template (1FL2 / XHC1) the red line indicates the *in silico* peptide junction:

```
P1;FLXH
structureX:FLXH:    :A: :A:undefined:undefined:-1.00:-1.00
----AYDVLIVGSGPAGAAAAIYSARKG-----IRTGLMGERFGGQILDTVDIENYISVPKTEGQKLAGALKVHVD
EYDVDVIDSQSASKLIPAAVEGGLHQIET----------ASGAVLKARSIIVLAT|GARAR-EPQI
KGKEYLLTLRTIFDADRIKESIENS------------GEAIIIGGGFIGLELAGNLAEA-----
----------GYHVKLIHRGAMFLGL-DEELSNMIKDMLEETGVKFFLNSELLEANEEGVLTNS-GFIEGKVKIC
AIGIVPNVDLARRSGIHTGR--GILIDDNFRTS-AKDVYAIGDCAEYSGIIAGT---AKAAMEQARVL
ADIL-----KGEPRRYNFKFRSTVFKFGKLQIAIIGNTKGEGKWIEDNTKVFYIGAVVFNDIRKATKLE*

>P1;ecCx
sequence:ecCx:     : :     : ::: 0.00: 0.00
LTTPLKKIVIVGGGAGGLEMATQLGHKLGRKKKAKITLVDRNHSHL--WKPLLHEVATGSLDEGVDALSYLAHARN
HGF-QFQLG-SVIDI-DREAK---TITIAELRDEKGELLVPERKIAYDTLVMALGSTSN-DFNTP
GVKENCIFLDNPHQARRFHQEMLNLFLKYSANLGANGKVNIAIVGGGATGVELSAELHNAVKQLH
SYGYKGLTNEALNVTLVEAGERILPALPPRISAAAHNELTKLGVRVLTQTMVTSADEGGLHTKDGEYIEADLMVW
AAGIKAPDFLKDIGGLETNRINQLVVEPTLQTTRDPDIYAIGDCASCPRPEGGFVPPRAQAAHQMATC
AMNNILAQMNGKPLKNYQYKDHGSLVSLSNFSTVGSLMGNLTRGSM*
```

In the final *E. coli* models alpha helix constraints were applied by deriving `ecmodel` from `automodel` as shown here for the second model:

```
from modeller.automodel import *

env = environ()
env.io.atom_files_directory = './:../../pdb_files/'

class ecmodel(automodel):
   def special_restraints(self,aln):
     rsr = self.restraints
     rsr.make(aln, restraint_type='ALPHA', residue_ids=('330', 348'),
              spline_on_site=False)
     rsr.make(aln, restraint_type='ALPHA', residue_ids=('180', '192'),
              spline_on_site=False)
     rsr.make(aln, restraint_type='ALPHA', residue_ids=('194', '202'),
              spline_on_site=False)
     rsr.make(aln, restraint_type='ALPHA', residue_ids=('144', '157'),
              spline_on_site=False)

a = ecmodel(env, alnfile='ecFXNB8.ali',
            knowns='FXNB', sequence='ecCx')
a.starting_model = 1
a.ending_model = 10
a.make()
```

The final template, FXNB, was composed of 6 peptide fragments and aligned as defined:

```
>P1;FXNB
structureX:FXNB: :A:  :A:undefined:undefined:-1.00:-1.00
----AYDVLIVGSGPAGAAAAIYSARKG-----
IRTGLMGERFGGQILDTVDIENYISVPKTEGQKLAGALKVHVD
EYDVDVIDSQSASKLIPAAVEGGLHQIETSLTAGDAYEQAAPTGEKAVLKARSIIVATGARAREPQI
KGKEYLLTLRTIFDADRIKESIENS--------------GEAIIIGGGFIGLELAGNLAEA-----
----------GYHVKLIHRGAMFLGL-DEELSNMIKDMLEETGVKFFLNSELLEANEEGVLTNS-
GFIEGKVKIC
AIGIVPNVDLARRSGIHTGR--GILIDDNFRTS-AKDVYAIGDATL
IKYN-PA-DTEV-NIAL-AKAAMEQARVLADIL----
-----KGEPRRYNFKFRSTVFKFGKLQIAIIGNTKGEGKWIEDNTKVFYIGAVVFNDIRKATKLE*

>P1;ecCx
sequence:ecCx:     : :     : ::: 0.00: 0.00
LTTPLKKIVIVGGGAGGLEMATQLGHKLGRKKKAKITLVDRNHSHLWKPLLHEVA--
TGSLDEGVDALSYLAHARN
HGF-QFQLG-SVIDI-DREAK---TITIAELRDEKGE----LLVPERKIAYDTLVMALGSTSNDFNT
PGVKENCIFLDNPHQARRFHQEMLNLFLKYSANLGANGKVNIAIVGGGATGVELSAELHNAVKQLH
SYGYKGLTNEALNVTLVEAGERILPALPPRISAAAHNELTKLGVRVLTQTMVTSADEGGLHTKDGEYIEA
DLMVW
AAGIKAPDFLKDIGGLETNRINQLVVEPTLQTTRDPDIYAIGDCAS
CPRP-----EGG-FVPPR-AQAAHQMATCAMNNILAQ
MNGKP-----LKNYQYKDHGSLVSLSNFSTVGSLMGNLTRGSM*
```

Loops were modelled using deriving a class from the `loopmodel`:

```
# Loop refinement of ecCx7

from modeller.automodel import *

log.verbose()
env = environ()

# directories for input atom files
env.io.atom_files_directory = './:../../../pdb_files'

class ecLoop(loopmodel):
    #  pick residues to be refined by loop modeling
    def select_loop_atoms(self):
        # 10 residue insertion
        self.pick_atoms(selection_segment=('158:', '170:'),
                        selection_status='INITIALIZE')

m = ecLoop(env,
           inimodel='ecCx8-6.pdb', # initial model of the target
           sequence='ecCx')          # code of the target

m.loop.starting_model= 1           # index of the first loop model
m.loop.ending_model  = 10          # index of the last loop model
m.loop.md_level = refine.very_fast  # loop refinement method

m.make()
```

An assessment of the quality of different models (in terms of steric impingements, distorted bond angles etc.) was made using a DOPE score. The log file for this python script reports these scores:

```
from modeller.automodel import *    # Load the automodel class

log.verbose()    # request verbose output
env = environ()  # create a new MODELLER environment to build this
model in
env.libs.topology.read(file='$(LIB)/top_heav.lib') # read topology
env.libs.parameters.read(file='$(LIB)/par.lib') # read parameters

# directories for input atom files
env.io.atom_files_directory = './:..'


def dope_profile(mdl, file):
    # DOPE energy parameters
    edat = energy_data(contact_shell=15.0, dynamic_modeller=True,
                       dynamic_lennard=False, dynamic_sphere=False,
                       excl_local=(False, False, False, False))
    # DOPE group restraints
    oldgprsr = mdl.group_restraints
    if not model.dope_restraints:
        model.dope_restraints = \
            group_restraints(classes='${LIB}/atmcls-mf.lib',
                             parameters='${LIB}/dist-mf.lib')
    mdl.group_restraints = model.dope_restraints
    molpdf = mdl.energy(output='SHORT', file=file,
                        edat=edat, residue_span_range=(1, 9999),
                        normalize_profile=True, smoothing_window=10)
    mdl.group_restraints = oldgprsr
    return molpdf

for i in range(9):
    # read model file
    code = "ecCx"
    num  = str(i+1)
    mdl = model(env)
    mdl.read(file='ecCx.BL000'+num+'0001.pdb')
    aln = alignment(env)
    # generate topology
    aln.append_model(mdl, atom_files='ecCx.BL000'+num+'0001.pdb',
align_codes=code)
    aln.append_model(mdl, atom_files='ecCx.BL000'+num+'0001.pdb',
align_codes=code+'-ini')
    mdl.generate_topology(aln, sequence=code+'-ini')
    mdl.transfer_xyz(aln)

    mdl.assess_dope(output='ENERGY_PROFILE NO_REPORT',
                    file='ecCx'+num+'.profile',
                    normalize_profile=True, smoothing_window=15)
```

Modelling of Arabidopsis NDH2s was achieved by manually aligning sequences against the *E. coli* sequence within a file that retained alignment with FXNB. Thus FXNB was used as the template to avoid cumulative errors arising from the use of a 'second generation' template:

```
>P1;FXNB
structureX:FXNB: :A:  :A:undefined:undefined:-1.00:-1.00
----AYDVLIVGSGPAGAAAAIYSARKG-----IRTGLMG-ERFGGQILDTVDI---
ENYISVPKT--EGQKLAGALKVHVD
EYDVDVIDSQSASKLIPAAVEGGLHQIETSLTAGDAYEQAAPTGEKAVLKARSIIVATGARAREPQ
I
KGKEYLLTLRTIFDADRIKESIENS-----------------GEAIIIGGGFIGLELAGNLAEA-
----
----------GYHVKLIHRGAMFLGL-DEELSNMIKDMLEETGVKFFLNSELLEANEEGVLTNS-
GFIEGKVKIC
AIGIVPNVDLARRSGIHTGR--GILIDDNFRTS-AKDVYAIGDATL
IKYN-PA-DTEV-NIAL-AKAAMEQARVLADIL----
------
KGEPRRYNFKFRSTVFKFGKLQIAIIGNTKGEGKWIEDNTKVFYIGAVVFNDIRKATKLE*

>P1;ecCx
sequence:ecCx:     : :     : ::: 0.00: 0.00
LTTPLKKIVIVGGGAGGLEMATQLGHKLGRKKKAKITLVD-RNHSHLWKPLLHE---VA--
TGSLD--EGVDALSYLAHARN
HGF-QFQLG-SVIDI-DREAK---TITIAELRDEKGE----
LLVPERKIAYDTLVMALGSTSNDFNT
PGVKENCIFLDNPHQARRFHQEMLNLFLKYSANLGANG----
KVNIAIVGGGATGVELSAELHNAVKQLH
SYGYKGLTNEALNVTLVEAGERILPALPPRISAAAHNELTKLGVRVLTQTMVTSADEGGLHTKDGE
YIEADLMVW
AAGIKAPDFLKDIGGLETNRINQLVVEPTLQTTRDPDIYAIGDCAS
CPRP-----EGG-FVPPR-AQAAHQMATCAMNNILAQ
MNGKP------LKNYQYKDHGSLVSLSNFSTVGSLMGNLTRGSM*

>P1;at1Cx
sequence:at1Cx:     : :     : ::: 0.00: 0.00
---EKPRVLVLGSGWAGCRVLKGIDTSI-----YDVVCVSPRNH-MVFTPLLASTC-
VGTLEFRSVA--EPISRIQPAISRE
PGS-YYFLA-NCSKL-DADN---HEVHCETVTEGSST----
LKPWKFKIAYDKLVLACGAEASTFGI
NGVLENAIFLREVHHAQEIRRKLLLNLMLSEVPGIGEDEKKRLLHCVVVGGGPTGVEFSGELSDFI
MKDV
RQRYSH-VKDDIRVTLIEARD-ILSSFDDRLRHYAIKQLNKSGVKLVRGIVKEVKPQKLIL-
DDGTEVPYGPLVW
STGVGPSSFVRSLDFPKDP-GGRIGIDEWMRVPSVQDVFAIGDCSG
YLEST----GKS-TLPAL-AQVAEREGKYLANLFNVM
GKAGG------GRANSAKEMELGEPFVYKHLGSMATIGRYKAL*
# (full) +VDLRESKEGKGIS
```

Chapter 3. NDH2 Sequence Analysis

3.1 *Arabidopsis* and other higher plant NDH2 genes.

3.1.1 *Arabidopsis thaliana*

Deposition of the 92%-completed genome sequence in 2000, including all supposed protein-coding regions, elevated *Arabidopsis thaliana* to an unique position among *viridiplantae* species for genetic-based research (The Arabidopsis Genome Initiative, 2000). It is thus ideally suited to higher plant NDH2 gene family examination. The protein sequence for At1g07180 (**nda1**), the most abundant NDH2 species, was obtained from an *Arabidopsis* genome website[1]. This was used as the query sequence for TBLASTN homology search (Altschul *et al.*, 1997) of this and the Genbank EST and NR databases[2], the latter species-restricted to *Arabidopsis thaliana* matches.
Nine distinct genes were identified in the genome as exhibiting high homology to **nda1** (including **nda1** itself), each with evidence of transcript expression (cDNA and EST records). These included the seven previously identified NDH2 genes (Michalecka *et al.*, 2003) and additionally 5g22140 and 3g44190. Each of these was then used as a query sequence for a second round of TBLASTN searching, which failed to elicit further homologous genes.

The nine protein sequences were submitted to a Mitoprot II server[3] for identification of putative mitochondrial targeting signals and cleavage sites for the mature proteins (fig 3.1). Individual alignments of transcripts against genomic data were then assembled for each gene to determine their intron/exon configurations (Supplementary Data S.1.2 to S.1.10), and finally an assembly of aligned *Arabidopsis thaliana* genes was constructed, annotated with this information (fig 3.2, and S.1.1).

[1] http://mips.gsf.de/proj/thal
[2] http://www.ncbi.nlm.nih.gov
[3] http://mips.gsf.de/cgi-bin/proj/medgen/mitofilter

Gene		probability	length	cleavage site
1g07180	nda1	0.7073	35	MLWIKNLARISQTTSSSVGNVFRNPESYTLSSRF|CTALQKQQVTDTVQAK
2g29990	nda2	0.9560	38	MFMIKNLTRISPNTSSIITRFRNSGSSSLSYTLASRF|CTAQETQIQSPAKIPN
4g28220	ndb1	0.7457	14	MTLLSSLGRASRS|APLASKLLLLGTLSGGSIVAYADANE
4g05020	ndb2	0.4655	12	MRNFSVFERFS|KAFKDHPSLTRILVVSTISGGGLIAYSEANASYGANGG
4g21490	ndb3	0.1240	12	MRPFAYFERLS|QAFHDYPSLSKILVVSTISGGGLIVYSEANPS
2g20800	ndb4	0.6340	11	MSFHSFYQRA|SSLFKAYPSTSKILLLSTFSGGGGV
5g08740	ndc1	0.9301	54	MAVLSSVSSLIPFSY...SSRWNSTRNSPMLYLSRA|VTNNSGTTEISDNE
5g22140		0.0392	-	MEGIESGSKQGKRVVVIGGGIAGSLA
3g44190		0.0406	-	MEKTESVSGKGKRVIVIGGGIAGSLA

Fig 3.1 Mitoprot II predicted targeting signals and cleavage sites

probability (of mitochondrial import): 0.0 (min) - 1.0 (max);
length: predicted cleaved signal peptide length.
BLUE: cleaved peptide, GREEN: mature protein, |: cleavage site.
In addition to the seven previously identified *Arabidopsis* genes, two putative genes identified by sequence homology (5g22140 and 3g44190) were submitted for Mitoprot II analysis.

From examination of this it is evident that all nine genes contain the pair of nucleotide binding folds (green boxes in fig 3.2) characteristic of FAD/NAD-linked reductases (SCOP) [FAD/NAD(P)H-binding Rossmann fold (PFAM)]. There is also significant homology across all genes at two further loci (Regions 1 and 2); however a third locus (Region 3) is clearly common to **nda, ndb** and **ndc** genes but finds no sequence similarity in 5g22140 or 3g44190. For this reason, and the absence of any putative mitochondrial targeting signal, it may be assumed that the latter are not NDH2 genes but possibly cytoplasmic FAD/NAD(P)H oxidoreductases using a second substrate other than quinone. (**ndb3** was predicted as mitochondrial rather than chloroplast directed, but with a lower level of certainty.)

The ~84-residue inserts in the **ndb** gene sequences (yellow box) were submitted to the MotifScan server[1] which confirmed the probable presence of an adjacent pair of calcium-binding EF-hand motifs[2] as previously reported (Michalecka *et al.*, 2003). The sequence homology of **ndc** to **nda** and **ndb** is poor (other than at the regions identified above) and no experimental evidence has been published to verify an NADH-quinone oxidoreductase activity. Modest sequence homology alone is not sufficient to

[1] http://myhits.isb-sib.ch/cgi-bin/motif_scan
[2] This motif may also bind other cations such as magnesium

determine that it is also a member of the NDH2 family; however its homology to other known prokaryotic NDH2s (below) and structural analysis (Chapter 5) provide compelling supporting evidence.

Compared to the three known NDH2s of *S. cerevisiae* and the (generally) single enzymes expressed in prokaryotic species, the seven genes of *Arabidopsis* constitute a substantial repertoire.

3.1.2 NDH2 in other higher plant species

The *Arabidopsis* **nda1**, **nda2** and **ndc** protein sequences were submitted to NCBI for TBLASTN search of the Genbank NR and EST databases to find homologous entries from other higher plant species.

Two rice **nda** genes were identified (*Oryza sativa,* AP003705 and AK069755, among others[1]), one for potato (*Solanum tuberosum* STU245861), one for tomato (*Lycopersicon esculentum,* BT013064) and a partial sequence for pea (*Pisum sativum,* AAO27256) (S.5.1). Three rice **ndb** genes were identified (AK102878, AK10088 and AK072710), one complete and one partial for maize (*Zea mays,* AY106876 and AY107518), and one each for potato (STU245862) and pea (BT013752) (S.5.2).Finally there is a single complete rice **ndc** gene (AK064752) and a partial EST for rape (*Brassica napus,* CD837336) (S.5.3). An analysis of *viridiplantae* intron structure and N-terminal signal sequences is presented below (section 3.5).

While it is possible that *Arabidopsis* is unique in possessing seven NDH2 genes it is more likely that the paucity of database depositions reflects the current state of *viridiplantae* sequencing: the genome sequencing project for rice, for example, is relatively advanced and six NDH2s are unambiguously identifiable.

[1] Both requiring manual editing due to incorrect intron prediction. This is not uncommon but can be corrected unambiguously when compared with known mature transcripts from other species, in this case *Arabidopsis* nda1.

Replace page with fig 3.2

3.2 NDH2 genes in all species

In order to collate putative NDH2 sequences across all species an exhaustive iterative method was adopted. Using *E. coli* NDH2 and *Arabidopsis* **nda1**, **ndb1** and **ndc** protein sequences as initial seeds the Genbank NR and EST databases were searched for homologous nucleotide sequences using TBLASTN. For example the initial search using **ndc** yielded the NDH2 gene from *Anabaena variabilis*; the latter was translated (Appendix A.3) and the protein sequence used in a subsequent search. This technique was continued until the gene mining process was terminated, when results contained duplications of NDH2s recovered by a different pathway (and by disregarding sequences which were self-evidently not NDH2 family members). Protein databases were not interrogated as there are almost invariably complementary nucleotide sequences in the NR (nucleotide) database, while the latter often harbour undiscovered/undocumented ORFs.

The principle challenge was to determine which proteins are truly NDH2s, a task that is compounded by the similarity of these enzymes to other FAD/NADH oxidoreductases, and the absence of an established distinguishing motif(s). Following the inspection of numerous candidates, selection was made on the basis of homology at 'Conserved Region 3' (fig 3.2). Many additional sequences were discovered exhibiting reasonable overall homology (and particularly at the nucleotide binding sites), but which bore no similarity at this locus.

As the purpose of this exercise was to identify conserved residues and motifs in NDH2 subfamily groups, rather than establish phylogenic relationships, a manual assembly of subfamily groups was undertaken rather than employing automated techniques (such as PAUP[1]). Although some subjectivity is inevitable, putative NDH2s were initially partitioned, by sequence homology, into one of ten groups (table 3.1). Genes were discovered in bacteria, fungi, plant and protozoan species, but no *animalia* putative NDH2s were discovered (as was expected).

[1] A phylogenic analysis program

Table 3.1 Classification of NDH2s

Table illustrating the classification used in this analysis.
Two structural variations from the canonical sequence may be determined:
1) the incorporation of putative EF-hand motifs (calcium-binding) in some fungal and plant genes.
2) five genes were identified homologous to that of *Acidianus ambivalens* which has been experimentally shown to include a covalently-bound FMN prosthetic group (in place of FAD). The sequences of this group are quite distinct from the canonical NADH.
features:

M : has covalently-bound FMN in place of FAD
EF: incorporates a putative calcium-binding ('EF-hand') domain

			feature	comments
1	Bacteria	Type 1		*E. coli* homologous enzymes
2		Type 2		other bacteria (including *archea*)
3		Cyanobacteria		cyano and homologous bacteria
4		FMN containing	M	covalently bound FMN (no FAD)
5	Fungi	Type 1		
6		Type 1a		minor sequence differences from Type 1
7		Type 2	EF	includes EF motif
8	Plant	nda		
9		ndb	EF	includes EF motif
10		ndc		closest homology to cyanobacteria

In the following sections a number of putative NDH2 genes (and their accession numbers) are provided for each of the groups shown in table 3.1. Also shown are alignments of selected loci from these genes, specifically: the N-terminus (including the first nucleotide binding region), the second nucleotide binding region, "Conserved Region 3" (see fig 3.2), and the C-terminus. The significance of "Conserved Region 3" as an NDH2-specific locus, first suggested by the analysis in fig 3.2, is confirmed by structural analysis (Chapter 5), as is the region following the first nucleotide binding site, and thus these are included in the following figures. The remaining loci are included to illustrate homology and variation across the groups. Full-sequence alignments are provided in the Supplementary Data.

3.2.1 Bacterial NDH2s

TBLASTN searching of bacterial nucleotide databases yielded over 120 putative NDH2 genes (including six of the *Acidianus ambivalens* FMN-containing type). These are partitioned here, according to sequence homology, into four groups: 1) those homologous to the *E. coli* NDH2, 2) diverse bacterial species, including the *archea*, 3) cyanobacteria and homologous species, and 4) FMN-containing NDH2s.

3.2.1.1 Bacteria Type 1

The *E. coli* NDH2 enzyme has been characterised and its gene unambiguously identified. Some 40 putative bacterial NDH2 genes may be grouped (table 3.2) on the basis of homology to the *E. coli* enzyme - though this group may alternatively be interpreted as a subgroup (albeit well represented in the databases) of the Type 2 bacterial NDH2s. Homology within this group is particularly high including many which differ in nucleotide but not amino acid sequence (fig 3.3, S.3.1).

The genomes of most species in this group encode a single NDH2, but there are exceptions: two homologous genes are present in *Vibrio fischeri, Ralstonia eutropha* and *Photobacterium profundum*, and three in *Burkholderia cepacia* Sp.383. It is not, however, possible to identify conserved residues which would permit segregation of these into (potentially functionally distinct) subgroups.

Near the N-terminus is the first of the recognised nucleotide binding fold, GxGxxG, which, in NDH2s, has been shown to interact with the bound FAD prosthetic group (Schmid & Gerloff, 2004). In all putative NDH2s this locus is preceded by four small residues (from I, V, L and A) (fig 3.3, and see also figures for other groups). In the case of enzymes of this group there are virtually no further residues between this and the initiation codon.

All putative NDH2s share a homologous second nucleotide binding site (for the NADH substrate); the consensus sequences for these loci in this group are **GGGAGG** and **GGGATG** respectively (slight variations of these characterise other groups' loci). The consensus motif at Region 3 is **PPRAQ**. The C-termini are generally conserved in both length and the final three residues (**KLH**), which are unique to this group.

Untypically, 13 of these genes (including that of *E. coli* itself) do not encode **ATG** at the translation start but **TTG** or **GTG**.

3.2.1.2 Bacteria Type 2

Group 2 bacterial NDH2s in this classification (table 3.3) contain those that are neither closely homologous to Type 1 (*E. coli*-like) not the cyanobacteria. There is greater diversity here in both sequence length and homology; in particular, there is not the consistency in C-terminus length (found in the other groups) so this locus has been omitted from the alignments (fig 3.4, see Supplementary data S.3.2 for full sequences).

Seven of these putative NDH2 genes (the first seven shown in fig 3.4) are distinct from the others in that they have an additional C-terminal domain (or domains) of some 250 residues. Whereas a few group 1 species appear to express two NDH2s (table 3.2), multiple NDH2s are the norm rather than the exception here, with, for example, three putative genes in some *Bacillus* species.

Table 3.2 Bacterial Type 1 NDH2s

NDH2s homologous to those of the *E. coli* NDH2 (biochemically verified, shown in **bold**) including genome accession numbers and number of putative genes.

Species	accession no.	NDH2s
Acinetobacter sp. ADP1	CR543861	1
Azotobacter vinelandii	AF346487	1
Bradyrhizobium japonicum	BA000040	1
Burkholderia cepacia	CP000152	3
Burkholderia xenovorans	CP000271	1
Chromohalobacter salexigens	EAM22718	1
Erwinia carotovora	BX950851	1
Escherichia coli	**AP009048**	**1**
Haemophilus ducreyi	AE017143	1
Haemophilus influenzae	CP000057	1
Hahella chejuensis	CP000155	1
Idiomarina loihiensis	AE017340	1
Mannheimia succiniciproducens	AE016827	1
Methylobacillus flagellatus	CP000284	1
Methylococcus capsulatus	AE017282	1
Photobacterium profundum	CR378670	2
Photorhabdus luminescens	BX571868	1
Pseudoalteromonas haloplanktis	CR954247	1
Pseudomonas aeruginosa	AE004867	1
Pseudomonas fluorescens	CP000076	1
Pseudomonas putida	AE015451	1
Pseudomonas syringae	CP000075	1
Pseudomonas xiamenensis	AF348165	1
Ralstonia eutropha	AY305378	2
Ralstonia solanacearum	AL646069	1
Rhodopseudomonas palustris	CP000250	1
Salmonella enterica	AE014613	1
Salmonella typhimurium	AE008752	1
Shigella boydii	CP000036	1
Shigella dysenteriae	CP000034	1
Shigella flexneri	AE005674	1
Shigella sonnei	CP000038	1
Sodalis glossinidius	AP008232	1
Thiobacillus denitrificans	CP000116	1
Yersinia pestis	AE017135	1
Yersinia pseudotuberculosis	BX936398	1
Vibrio parahaemolyticus	BA000031	1
Vibrio cholerae	AE004264	1
Vibrio fischeri	CP000020	2
Vibrio vulnificus	BA000037	1
Wigglesworthia glossinidia	BA000021	1

```
     1          nuc-binding site 1                                  *          59
ec   LTTPLKKIVIVGGGAGGLEMATQLGHKLGRK-----KKAKITLVDRNHSHLWKPLLHEVATGSL
sf   LTTPLKKIVIVGGGAGGLEMATQLGHKLGRK-----KKAKITLVDRNHSHLWKPLLHEVATGSL
se   LTTPLKKIVIVGGGAGGLEMATQLGHKLGRK-----KKAKITLVDRNHSHLWKPLLHEVATGSL
st   LTTPLKKIVIVGGGAGGLEMATQLGHKLGRK-----KKAKITLVDRNHSHLWKPLLHEVATGSL
sg   LTTPLKNIVIIGGGAGGLELATSLGRKLGRR-----KRAKVTLVDRNHSHLWKPLLHEVATGSL
yp   LTTLKKKIVIIGGGAGGLELATSLGHKLGRS-----KKADIVLVDRNHSHLWKPLLHEVATGSL
eca  LTSPTKKIVIVGGGAGGLELATSLGHTLGRK-----KKADITLVDRNHSHLWKPLLHEVATGSL
pl   MTTPKKRIVIIGGGAGGLELATSLGHKLGRK-----KKAEIVLVDRNQSHLWKPLLHEVATGSL
vf1      MTKIVIVGGGAGGLELATKFGRTLGRK-----GRAEVTLVDRKASHLWKPLLHEVATGSL
pp1      VTRIIVVGGGAGGLELATKLGRTLGRK-----GRAQVTLVDRKASHLWKPLLHEVATGSM
vv       VTRIIVVGGGAGGLELATKLGRTLGRK-----RRAQITLVDRKASHLWKPLLHEVATGSL
vp       VTRIIVVGGGAGGLELATKLGRTLGRK-----NRAQITLVDRKASHLWKPLLHEVATGSL
vc   RIVVVTRIIVVGGGAGGLELVTKLGRTLGRK-----GRANVTLVDRNSSHLWKPLLHEVATGSL
hd     MAKETIIIVGGGAGGLELATYLGNKLSRK-----NKAHVLLIDRNTTHLWKPLLHEVATGSL
ms       MKNIVIVGGGAGGLELATYLGNNLGKK-----QRANVVLVDRNQTHLWKPLLHEVATGVL
hi       MKNVVIVGGGAGGIELATFLGNKLGRQ-----KQAKVTLVDRNATHLWKPLLHEIATGVM
wg   KNKNLKKIIIVGGGAGGLELVTKLGNNLGKK-----KLAKIILVDNNYIHLWKPWLHKFASGSL
hc     MSIPKIVVVGGGAGGLELVTRLGNALGKK-----KRAVVHLVDCNPTHLWKPLLHEVATGSM
il      MQERIVIVGGGAGGLELATMLGRKLGRK-----KKAKIMLIDRNSTHLWKPLLHEVASGAL
mc   NEQPLPRILIIGGGAGGLELATKLGRTLGKK-----GKAQVTLIDAHPTHIWKPLLHEVAAGTM
av       MTHRIVIVGGGAGGVELATRLGKTMGRN-----FQAKITLVDANMTHLWKPLLHEVAAGSL
psp      MTHRIVIVGGGAGGLELATRLGKSLGKR-----KQAEITLVDTNLTHIWKPLLHEVAAGSL
psf      MTHRIVIVGGGAGGLELATRLGKTLGKR-----GTASVMLVDANLTHIWKPLLHEVAAGSL
pss      MTHRIVIVGGGAGGLELATSLGKTLGKK-----GTASVTLVDANLTHIWKPLLHEVAAGSL
psa      MSHRIVIVGGGAGGLELATRLGRTLGKR-----GKAQVTLVDANLTHIWKPLLHEVAAGSL
bj   AQAGKPRVVIVGGGAGGLELATRLGDKYGRK-----GRLDVTLIERNRTHVWKPKLHEIAAGSM
rp   MSDKPHRIVIVGGGAGGLELATALGNKYGKS-----DAVRVTLVDRSRTHIWKPLLHAIAAGSL
re2  PADSQHQIVIVGGGAGGLELATALGNKLGKR-----GLAAVTLIDKTRSHIWKPKLHEIAAGSM
re1  ANGRAHRIIVVGGGAGGLELVTRLGDKLGKK-----GTAQVVLVDRLPTHIWKPLLHEVAAGSM
cs     MSTPRIVVVGGGAGGLELVTRLGRKLGKR-----QRAEIVLVDRNATHIWKPLLHEVATGVL
bc1  KGEQMHRIIIVGGGAGGLELATRLGDRYGAR-GNRSARALVTLVDRNPTHIWKPLLHEVAAGSM
bx       MHRFIVVGGGAGGLELATRLGDRYAPHKNKGAVRAQVTLVDRNPTHIWKPLLHEVAAGSM
mf   TNSDLHRVVIVGGGAGGLELATRLGRTLGRK-----KKAHITLIDCTRTHVWKPLLHEIAAGSM
td    VTEKHEIVIVGGGAGGLELATRLGDKLGGK-----KRANITLIDASPSHLWKPLLYEVAAGSL
aS   QTSNLHRIVIVGGGAGGLELATRLGNQLGKH-----QHAEIILVDAALTHIWKPLLHEVAAGTL
bc2   MTTPTRIVIVGGGIAGLQLATRLGERLGRS-----GRAQVTIVDRSPTHIWKPMLHTIAAGTR
bc3  RAPRLPRIVIVGGGAGGLHLATRLGDTVGRR-----GQAEVVLVDRYPTHFWKPLLHEAASGHR
vf2      MSRIIVVGGGAAGLELSTLLTKST-RK------EDEIILVEPETHHYWKPRLHEIAAGTF
pp2      MSKIVVVGGGAAGLELVTRLGRSFGRR-----DTHEIILVEPSTHHYWKPRLHEIAAGTF
```

Fig 3.3 (1/2) Alignment of Bacterial Type 1 sequences

N-terminus region including first nucleotide binding region

(top) *E. coli* residue number.
*: Trp residue discussed in Chapter 6.
Blue numbers distinguish multiple NDH2 genes within a species.
Species abbreviations (for accession numbers see Table 3.1): ec *Escherichia coli*; sf *Shigella flexneri*; se *Salmonella enterica*; st *Salmonella typhimurium*; sg *Sodalis glossinidius*; yp *Yersinia pestis*; eca *Erwinia carotovora*; p *Photorhabdus luminescens*; vf *Vibrio fischeri*; pp *Photobacterium profundum*; vv *Vibrio vulnificus*; vp *Vibrio parahaemolyticus*; vc *Vibrio cholerae*; hd *Haemophilus ducreyi*; ms *Mannheimia succiniciproducens*; hi *Haemophilus influenzae*; wg *Wigglesworthia glossinidia*; hc *Hahella chejuensis*; il *Idiomarina loihiensis*; mc *Methylococcus capsulatus*; av *Azotobacter vinelandii*; psp *Pseudomonas putida*; psf *Pseudomonas fluorescens*; pss *Pseudomonas syringae*; psa *Pseudomonas aeruginosa*; bj *Bradyrhizobium japonicum*; rp *Rhodopseudomonas palustris*; re *Ralstonia eutropha*; cs *Chromohalobacter salexigens*; bc *Burkholderia cepacia*; bx *Burkholderia xenovorans*; mf *Methylobacillus flagellatus*; td *Thiobacillus denitrificans*; as *Acinetobacter sp. ADP1*.

	nuc- binding site 2		' region 3'		C terminus
	173	185	327	339	434
ec	..IAIVGGGATGVEL..		..VPPRAQAAHQMAT..		..GSINRVIRPRLKLH*
sf	..IAIVGGGATGVEL..		..VPPRAQAAHQMAT..		..GSINRVIRPRLKLH*
se	..IAIVGGGATGVEL..		..VPPRAQAAHQMAT..		..GSINRVIRPRLKLH*
st	..IAIVGGGATGVEL..		..VPPRAQAAHQMAT..		..GSINRVIRPRLKLH*
sg	..IAIVGGGATGVEL..		..VPPRAQAAHQMAS..		..GRINRVIRPHLKLH*
yp	..IAIVGGGATGVEL..		..VPPRAQSAHQMAS..		..GGINRVIRPRLKLH*
eca	..IAIVGGGATGVEL..		..VPPRAQAAHQMAS..		..GGINRVIRPRLKLH*
pl	..IAIVGGGATGVEL..		..VPPRAQAAHQMAS..		..GGINRVIRPRLKLH*
vf1	..IAIVGAGATGVEL..		..VPPRAQAAHQMAS..		..GRINRVLRPNLKLH*
pp1	..IAIVGAGATGVEL..		..VPPRAQAAHQMAS..		..GRINRVLRPNLKLH*
vv	..IAIVGAGATGVEL..		..VPPRAQAAHQMAS..		..GRINRVLRPNLKLH*
vp	..IAIVGAGATGVEL..		..VPPRAQAAHQMAS..		..GRINRVLRPNLKLH*
vc	..IAIVGAGATGVEL..		..VPPRAQAAHQMAS..		..GRINRVLRPNLKLH*
hd	..IAIVGGGATGIEL..		..IPPRAQAAHQMAT..		..GQINRLLRPSMKLH*
ms	..IAIVGGGATGVEL..		..VPPRGQAANQMAT..		..GRLNRFIRPSLKLH*
hi	..IAIVGGGATGVEL..		..VPPRAQAAHQMAK..		..DKLNRYLKPRLKLH*
wg	..ISIVGGGATGVEL..		..VPPRAQAAHQMAR..		..TGIENIISPKLKLH*
hc	..IGIVGAGATGVEL..		..VPPRAQSAHQMAS..		..ARINKIIRPRLKLH*
il	..LAIVGGGATGVEL..		..VPPRAQSAHQMAE..		..DKLNHVVKPRLKLH*
mc	..VAIVGGGATGIEL..		..VPPRAQSAHQQAS..		..HFFSQKLRPKIKLH*
av	..VAIVGAGATGVEL..		..VPPRAQAAHQQAS..		..DRFRSSTEPRLKLH*
psp	..VAIVGAGATGVEL..		..VPPRAQAAHQQAS..		..SKIGRGTEPRLKLH*
psf	..VAIVGAGATGVEL..		..VPPRAQAAHQQAS..		..SKIGRGTEPRLKLH*
pss	..VAIVGAGATGVEL..		..VPPRAQAAHQQAS..		..SKIGKGTEPRMKLH*
psa	..IAIIGAGATGVEL..		..VPPRAQAAHQQAS..		..GRL-RSTEPRLKLH*
bj	..VAIIGAGATGVEL..		..VPPRAQAAHQQAS..		..RLITRRTEPHVKLH*
rp	..VAIIGAGATGTEL..		..VPPRAQAAHQEAE..		..GST-RPT-PTVKLH*
re2	..VAIIGAGATDVEL..		..VPPRAQAAHQQAS..		..RLITRRTEPHVKLH*
re1	..VAIIGAGATGVEL..		..VPPRAQAAHQQAT..		..HWLRSKTSPRVKLH*
cs	..VGLVGAGATGVEL..		..VPPRAQAAHQQAS..		..DGLNRYLRPRMKLH*
bc1	..VAIVGGGATGVEL..		..VPPRAQAAHQQAN..		..HWLRRTTLPRVKLH*
bx	..VAIVGGGATGVEL..		..VPPRAQAAHQQAS..		..HWLRRTTLPRVKLH*
mf	..VVIVGAGATGVEL..		..VPPRAQAAHQQAS..		..RLITRRTEARVKLH*
td	..VGIVGAGATGVEL..		..VPPRAQAAHQQAS..		..HFISTPTRARIKLH*
aS	..IAIIGAGATGVEL..		..VPPRAQVANQEAI..		..DLLSKGSSPQLKLH*
bc2	..VAIVGAGATGVEL..		..LPPTAQVATQQAE..		..ERINGCVQPRIRLS*
bc3	..INVIGAGATGVEL..		..LPPRAQVAYQQAV..		..HWLQSRNQPSIKLH*
vf2	..INIVGAGATGVEL..		..VPPKAQAANRAAV..		..KRLKAFFKPAGI*
pp2	..LSIVGAGATGVEL..		..VPPLAQAANQAAG..		..CALNPYYKGNIS*

Fig 3.3 (2/2) Alignment of Bacterial Type 1 sequences

Second nucleotide-binding site, "conserved region 3" and C terminus.

Table 3.3 Bacterial Type 2 NDH2s

Bacterial NDH2s other than *E. coli* homologous and cyanobateria, including genome accession numbers and number of putative genes.

Species	accession no.	NDH2s
Bacillus anthracis	NC_003995	3
Bacillus cereus	NC_004722	3
Bacillus halodurans	BA000004	2
Bacillus licheniformis	CP000002	3
Bacillus subtilis	NC_000964	3
Bacteroides fragilis	CR626927	1
Bacteroides thetaiotaomicron	AE015928	1
Clostridium thermocellum	NZ_AABG02000072	1
Deinococcus radiodurans	NC_001263	1
Enterococcus faecalis	AE016957	2
Geobacillus kaustophilus	BA000043	2
Geobacter sulfurreducens	AE017180	1
Heliobacillus mobilis	AY142850	1
Kuenenia stuttgartiensis	CT573072	1
Lactobacillus plantarum	AL935254	2
Lactococcus lactis	AE006316	2
Listeria innocua	AL596173	2
Listeria monocytogenes	AL591983	2
Nitrosospira multiformis	CP000103	1
Oceanobacillus iheyensis	NC_004193	1
Oenococcus oeni	NZ_AABJ02000001	1
Staphylococcus aureus	NC_003923	2
Streptococcus agalactiae	NC_004368	1
Thermoanaerobacter tengcongensis	NC_003869	1
Cenarchaeum symbiosum (archea)	DQ397575	1
Haloarcula marismortui (archea)	AY596297	1
Halobacterium sp. NRC-1 *(archea)*	AAG19331	1
Natronomonas pharaonis (archea)	CR936257	1
Sulfolobus acidocaldarius (archea)	CP000077	1
Sulfolobus solfataricus (archea)	AE006905	2
Sulfolobus tokodaii (archea)	NC_003106	1

```
        3          nuc-binding site 1                                    64
ll2    MTKKQIVVIGGGVAGVHATRELS--KKLKADADITLIDKHSYHTTMTQLHEVAAG-RVPFTT
li2    MPEKNIVLIGAGYAGVHAAKKLAKKYKKDKDVNITLIDRHSYHTMMTELHEVA-GGRVEPTA
lm2    MPEKNIVLIGAGYAGVHAAKKLAKKYKKDKDVNITLIDRHSYHTMMTELHEVA-GGRVEPTA
tt    MGNNKKIVIIGAGYGGLHAAKLLNKKLKNNPDIEVTLIDKKPYHTLLTDLHEVA-GSRIEPDS
ct     MAHKVLILGAGYAGIEAALTLHKKKRKKDDIEITIIDKNPYHTLLTELHEVA-GNRITEDA
ef2    MTKQNIVVVGAGYAGVSATKFLAKKFKKDTDVTITLIDRHSYHTMMTELHEVA-GGRVEPEA
lp2    MAKKNIVVVGAGFAGVYATKKLSKHFKKNADVEITLIDRHSYFTYMTELHEVA-TERVEPEH
ll1    MAKKKIVIVGAGFAGVSATRLLA--KNLKNDAEITVIDKHDFQTSMTQLHEVA-ASRVEADA
ks    MESKKNIVVVGAGYGGITSVLRLARLFRKHPEYQIHLIDRNPYHTLKTQLHEAA-VRKRE--V
bs1     MSKHIVILGAGYGGVLSAL-TVRKHYTKEQARVTVVNKYPTHQIITELHRLAAGNVSEKAV
bl1     MSKHIVILGAGYGGLLSAL-TVRKHYTKEEAKVTVVNQYPTHQIITELHRLAAGNVSEQAV
ba1     MSKQIVILGAGYGGLLAAL-NVRKYYSKSEAQVTVINQYPTHQIITELHRLAAGNVSEQAV
bc      MSKQIVILGAGYGGLLAAL-NVRKYYSKSEAQVTVINQYPTHQIITELHRLAAGNVSEQAV
dr        MKTLILGAGYAGLATT----TSLKPIPGLESLLIEQNAYHTFETRLHEAAAHNA---RI
lm1    MSKPKIVILGAGYGGLKTLR---KLQQRNLEAEIVLVNKNDYHHETTWLHEAAAGTIEPEKL
li1    MSKPKIVILGAGYGGLKTLR---KLQQRNLDAEIVLVNKNDYHHETTWLHEAAAGTIEPEKL
bl3    LNKPKVVVLGAGYGGLMTVT-RLVKKIGINEANITLVNKHNYHYETTWMHEASAGTLHHDRC
bs3  MALNKPKIVILGAGYGGLMTVT-RLTKYVGPNDADITLVNKHNYHYETTWMHEASAGTLHHDRC
bh1    LNKPNIVILGAGYGGIITAS-RLSKQLGHNDANITLVNKHSYHYQTTWLHEPAAGTIPAEKA
ba3   LVKTPKIVVLGAGYGGMITTV-RLQKALSVNEAEITLVNNNSYHYQATWLHESAAGTLQDEKI
gk1    VRKPNVVILGAGYGGLMTTV-RLQKLIGVNEANITLVNKHDYHYETTWLHEASAGTLHHDRV
sa1  MAQDRKKVLVLGAGYAGLQTVTKLQK-AISTEEAEITLINKNEYHYEATWLHEASAGTLNYEDV
oi     MSKPKVVVLGAGYAGLVATR-RLTQKLSADEAEIVLINKHNYHYESTWLHEVAAGTINPNQA
hm   MKAGKPRIVVLGAGYAGILTTR-RLQNLLSSDEAEIVLVNKHNYHYLTTWLHEVAAGTGDDDRI
ef1    MNKKHVVILGAGYAGLKTLR---ELQKGAKDVEITLVDRNNYHYEATDLHEVAAGTQPAEKI
sa       MKEILVLGAGYAGLKAVR---NLQKQSGDFHITLVDMNDYHYEATELHEVAAGSQPKEKI
lp1      MATTLVLGGGYAGMRAVK--FLQKSLPTEDEIILVDQTPTHTEKTNLHEVAAGTIAPDRI
oo       MKTIVVLGAGYAGLRVVR---ELVDHKANAKIVLINKNSYHYESTQLHEVAIGSKSPNDI
(ec)LTTPLKKIVIVGGGAGGLEMATQLGHKLGRKKKAKITLVDRNHSHLWKPLLHEVATGSLDEGVD
(mt)QPPRRHRVVIIGSGFGGLNAAK-----KLKRADVDIKLIARTTHHLFQPLLYQVATGIISEGEI
nm      MKTKIAIIGGGFAGLNLVK----HLAGKEEFDVTLVDMNNYHLFPPLLYQAATGFLDVSNI
bf   VKNDKKRVIIVGGGFGGLKLAN-----KLKKSGFQVVLVDKNNYHQFPPLIYQVASAGLEPSSI
bt   AKDSKKRVVIVGGGFGGLKLAN-----KLKKSGFQVVLIDKNNYHQFPPLIYQVASAGMEPTSI
gs       VKRVVIIGMGFGGIRAAR-----TLAQKGLDVVLVDRNNYHLFQPLLYQVATAGLEQESI
ba2      MKHLVILGGGYGGMRILQRLLPSNQLPDDVQVTLIDKVPYHCFKTEYYALVAGTIS---E
sa2      MKNLVLLGGGYGNMRIMSRILTTS-LPQDYTVTLVDRMPFHGLKPEFYALAAGTKS---D
gk2      MRQLVLLGGGYGNMRILQRLLPDELP-NDIHIILIDRVPYHCLKTEYYALAAGTVS---D
bl2      MRNLVLLGGGYGNMRIIHRLLPNQLP-DDVMITLIDRNPYHCLKTEYYALAAGTIS---D
bs2      MKKLVLIGGGYGNMRVLHRLLPNQLP-DDVSITLIDRNPYHCLKTEYYALAAGTIS---D
bh2      MKKLVVLGGGYGGMRILQRLLPNDLP-SDWEITLVDRLPYHCLKTEYYALAAGTAS---D
hs      MTTRVVVLGAGYAGAGAIAGLESELGPD--AELVWVADTDYHLVLHEAHR-VIRDPGVQEK
np      MTDDVVVLGSGYAGTGAINSLESELDGE--ADITWISDVDHHLVLHESHR-CIRDPDIQEK
hma     MTENVVVLGSGYAGAGAIKSFEDELDGQTDVDVTWISETDYHLVLHESHR-CIRDPSVQEN
cs   LQDKKKRILILGGGFAGVKCARELESTMGGDADVEITMVSEDNFLLFTPMLPQ--VASGMIETR
ss1    MQGERVVILGGGFAGIAAKLVY---------PNAILVDEKDFMVVTPRLVEVI--------E
st       MKVVILGGGFAGISAKLSY---------PNSILIDENDFVVNTPKLVEVIE-------N
sa       MRVVILGGGFAGLSALK---------TYRNSILIDEKDYFVLTHRLVDVV---------
ss2       MILILGGGFAGVSAYNQN--------KENSLVVDRKDYFLLTPWIIDFICGMKKLEDI
```

Fig 3.4 (1/2) Alignment of Bacterial Type 2 sequences

N-terminus region including first nucleotide binding region
(top) Residue numbered according to *E. coli* residue.
Also shown: *E. coli* ('ec', Type I) and *Mycobacterium tuberculosis* ('mt',Type 3)
Blue numbers distinguish multiple NDH2 genes within a species.

Species abbreviations (for accession numbers see Table 3.2): ll *Lactococcus lactis*, li *Listeria innocua*, lm *Listeria monocytogenes*, tt *Thermoanaerobacter tengcongensis*, ct *Clostridium thermocellum*, ef *Enterococcus faecalis*, lp *Lactobacillus plantarum*, ks *Kuenenia stuttgartiensis*, bs *Bacillus subtilis*, bl *Bacillus licheniformis*, ba *Bacillus anthracis*, bc *Bacillus cereus*, dr *Deinococcus radiodurans*, bh *Bacillus halodurans*, gk *Geobacillus kaustophilus*, sa *Staphylococcus aureus*, oi *Oceanobacillus iheyensis*, hm *Heliobacillus mobilis*, sa *Staphylococcus aureus*, oo *Oenococcus oeni*, nm *Nitrosospira multiformis*, bf *Bacteroides fragilis*, bt *Bacteroides thetaiotaomicron*, gs *Geobacter sulfurreducens*, hs *Halobacterium* sp. NRC-1, np *Natronomonas pharaonis*, hma *Haloarcula marismortui*, cs *Cenarchaeum symbiosum*, ss *Sulfolobus solfataricus*, st *Sulfolobus tokodaii*, sa *Sulfolobus acidocaldarius*..

	nuc- binding site 2		region 3	
	173	185	327	339
ll2	..ISVAGSGFTGIEM..		..TPQIVEAAEQTAH..	
li2	..FVVCGSGFTGIEM..		..TPQIVEGAEQTAL..	
lm2	..FVVCGSGFTGIEM..		..TPQIVEGAEQTAL..	
tt	..FVIAGAGFTGIET..		..IPQVVETALQSAE..	
ct	..FVVGGGGFTGVET..		..LPALVEAALQTGK..	
ef2	..FVVCGSGFTGIEM..		..TPQIVQAAEQTGH..	
lp2	..FTVCGSGFTGSEL..		..VPQIVQGAEETAA..	
ll1	..FSVAGSGFTGIEM..		..TPQIVEAAEQTGS..	
ks	..FVIGGGGLSGIEF..		..VPAAAQFALQQGR..	
bs1	..ILIGGGGLTGVEL..		..YPPTAQIAWQMGE..	
bl1	..ILIGGGGLTGVEL..		..YPPTAQIAWQMGE..	
ba1	..IVIGGGGLTGVEL..		..YPPTAQIAWQMGE..	
bc	..IVIGGGGLTGVEL..		..YPPTAQIAWQMGE..	
dr	..IVVGGAGLTGVEL..		..VPTTAQHAGQQGR..	
lm1	..IIVGGAGFTGIEF..		..FPPTAQIAMQQAD..	
li1	..IIVGGAGFTGIEF..		..FPPTAQIAMQQAD..	
bl3	..IVVGGAGFTGIEF..		..YPPTAQISMQQGE..	
bs3	..IVVGGAGFTGIEF..		..YPPTAQIAMQQGI..	
bh1	..FIVAGAGFTGIEF..		..YPPTAQIAIQMAE..	
ba3	..IVVGGAGFTGIEY..		..YPPTAQIAIQQGY..	
gk1	..IVVGGAGFTGIEF..		..YPPTAQIAMQEGQ..	
sa1	..ILVGGAGFTGVEF..		..LPTTAQIAMQQGE..	
oi	..ILVGGGGFTGIEF..		..YPPTGQLATQEGA..	
hm	..FIVGGAGFTGIEF..		..YPPTAQLAILQSD..	
ef1	..IVVCGAGFTGIEL..		..YPTTAQIALKMGA..	
sa	..LLVCGAGFTGIEL..		..FPTTAQIATRMGA..	
lp1	..IAVCGAGFTGIEL..		..YPTTAQIALAAGE..	
oo	..IAVCGAGFTSIEY..		..YPTTGQISVAQAT..	
(ec)	..IAIVGGGATGVEL..		..VPPRAQAAHQMAT..	
(mt)	..FTVVGAGPTGVEM..		..VPGVAQGAIQGAK..	
nm	..VVIAGGGPTGVEI..		..HPQLAQVALQQGR..	
bf	..VVVVGGGATGVEI..		..HPQLAQVAIQQGE..	
bt	..IVVVGGGATGVEV..		..HPQLAQVAIQQGE..	
gs	..FVIVGGGPTGVEF..		..LPMVAPVAMQMGI..	
ba2	..VAVVGAGLSGVEV..		..HAPSAQLAEGQGE..	
sa2	..VGIVGAGLSGIEL..		..HAPSAQLAEVQGD..	
gk2	..VGIVGAGLSGVEL..		..HSPSAQLAEAQAE..	
bl2	..VAIVGAGLSGVEL..		..HAPSAQLAEAQAE..	
bs2	..VAIVGAGLSGVEL..		..HAPSAQLAEAQAE..	
bh2	..VTIVGAGLSGVEL..		..HAPSAQLAEGQAE..	
hs	..VVIGGAGLSGIQS..		..APPTAQAAWQAAW..	
np	..VIVGGAGLSGIQS..		..APPTAQAAWQAAW..	
hma	..VVIGGAGLSGIQT..		..APPTAQAAWQAAW..	
cs	..FVIVGGGFAGIET..		..FPPTAQLAEAQGE..	
ss1	..VTILGGGALGVEL..		..IPQSAQVSLQAGL..	
st	..VTILGGGALGIEL..		..IPQSAQVASQAGS..	
sa	..VAVLGGGALSVEL..		..IPMSAQVAVQAGV..	
ss2	..ITIIGGGATGVEL..		..IPMTADIAVAAGV..	

Fig 3.4 (2/2) Alignment of Bacterial Type 2 sequences

Second nucleotide-binding site and "conserved region 3"

3.2.1.3 Cyanobacterial NDH2s

Using the *Arabidopsis* **ndc** as an initial query string yielded a number of *cyanobacteria* genes, in addition to those for rice and rape (S.5.3). Further homology searching yielded a distinct group containing all the cyanobacterial genes and a few others such as those from *Mycobacterium* species (table 3.4, S.3.3). These are quite distinct from the bacterial NDH2s listed above, bearing scant homology other than at the conserved regions.

As is the case with the type 2 species there is considerable diversity in the C-termini lengths and residues, and multiple putative NDH2 genes are again common, but, most remarkably, there appear to be as many as six in *Anabaena/Nostoc* species.

Table 3.4 Cyanobacterial NDH2s

Cyanobacterial NDH2s and homologous species.

Species	accession no.	NDH2s
Agrobacterium tumefaciens	AE009153	1
Anabaena variabilis	CP000117	6
Calothrix viguieri	CVPME131	1
Corynebacterium glutamicum	**BX927152**	**1**
Crocosphaera watsonii	NZ_AADV02000002	2
Gloeobacter violaceus	BA000045	2
Mycobacterium avium	AE016958	2
Mycobacterium smegmatis	AF038423	1
Mycobacterium tuberculosis	AE000516	2
Nocardia farcinica	AP006618	2
Nostoc sp. PCC 7120	BA000019	6
Prochlorococcus marinus	BX572099	1
Rhodopirellula baltica	BX294142	1
Synechococcus elongates PCC 7942	CP000100	3
Synechocystis sp. PCC 6803	BA000022	3
Solibacter usitatus	NZ_AAIA01000044	1
Trichodesmium erythraeum	NZ_AABK04000002	1

```
                1          nuc-binding site 1                          53
gv1      MLEGSFTASAMARIVIVGGGFAGLFTALGLEA-YPFKDERPEILLIDRSERFVFSPLLYE
se1            MVSANSAICILGGGFGGLYTALALAQQSWQGASRPPIHLIDRGDRFVFLPLLYE
ss1            MTDARPRICILGGGFGGLYTALRLGQLSWEGHTPPEIVLVDQRDRFLFAPFLYE
ns1            MTEQTKRIVILGGGFGGLYTALRVSQLPWETQQKPEIVLVDQSDRFLFSPLLYE
av1            MTEQTKRIVILGGGFGGLYTALRLSQLPWETQQKPEIVLVDQSDRFLFSPLLYE
se2         MIASTSIAKPTVIVGGGFVGLFCALHLRH----RHYPAPIILIDPKDRFIFRPLLFD
pm          MIPNPSNANAVVVVGGGFAGLTTALALSH----CQPRPPIVLIEPRQRFVFLPLLYE
ss2 -GVGWKTVAINSSDSATTVIIGGGFVGLFTALHLRH----HQHAGPIVLVEPQANFVFKPMLYE
ns2            MQQASHPTVILGGGFAGLFTALHLSQ----QNYSYPIILIEQRDRFSFKPLLYE
av2            MNSPIYQTVIVGGGFTGLFTALHLAH----EHYPRSVILIDRNERFCFKPLLYE
gv2          MPTMMPSDRTVILGGGFTGLFTALRLNR----RRYPHPVVLVDRSERFSFKPLLYE
mt1  MTLSSGEPSAVGGRHRVVIIGSGFGGLNAAKALKR------ADVDITLISKTTTHLFQPLLYQ
mt2  MSPQQEPTAQPPRRHRVVIIGSGFGGLNAAKKLKR------ADVDIKLIARTTHHLFQPLLYQ
ma1 -DRPPGTAEGTGMSRHRVVIIGSGFGGLTAAKALKR--VPEGTQVDITLISKTTTHLFQPLLYQ
ma2  MSPHSGSTAGPERRHQVVIIGSGFGGLNAAKKLKH------ANVDIKLIARTTHHLFQPLLYQ
ms       MSHPGATASDRHKVVIIGSGFGGLTAAKTLKR------ADVDVKLIARTTHHLFQPLLYQ
nf1       MSNRTGENRRHQVVVIGSGFGGLFGTKHLKR------ADVDVTLISKTSTHLFQPLLYQ
nf2  MDNVAVVEATGIEYSDVVIVGSGFGGLAAAKQLAK------SNVDYVLISSTPEHLFQPLLYQ
cg     MSVNPTRPEGGRHHVVVIGSGFGGLFAAKNLAK------ADVDVTLIDRTNHHLFQPLLYQ
av6              MRNRRVVIVGAGFGGLQAAQSLAN------SGADVLLIDRHNYHTFVPLLYQ
ns6              MMRNRRVVIVGAGFGGLQAAQSLAH------SGADVLLIDRHNYHTFVPLLYQ
rb          MSNATKQRPHVVVVGGGFAGLQATRDLRK------VDVSVTLLDRRNFHLFQPLLYQ
av5        MVDLHEKKAPHQVVIVGGGFGGLYTAKTLAK------ANVSVTLIDKRNFHLFQPLLYQ
ns5        MVDAHEKKAPHQVVIIGGGFGGLYTAKTLAT------ANVSVTLIDKRNFHLFQPLLYQ
av3           MVASLEPHQVVIVGGGFGGLYAAKALAK------ANVNVTLIDKRNFHLFQPLLYQ
ns3        MVASLENHQPHEVVIVGGGFGGLYAAKALAK------TNVNVTLIDKRNFHLFQPLLYQ
cv         MVASVEKNQRHQVVVIGGGFGGMYTAKGLAS------ANVDVTLIDKRNFHLFQPLLYQ
av4        MVEALDNNPPHKVVIVGGGFGGLYAAKALAK------AKVDVTLIDKRNFHLFQPLLYQ
ns4        MVEALDNNPPHKVVIVGGGFGGLYAAKALAK------AKVDVTLIDKRNFHLFQPLLYQ
se3           MISTVSKPRVVVIGGGFGGLYTALNLGK------TSVELTLIDKRNFHLFQPLLYQ
ss3         MNSPTSPRRPHVVIVGGGFAGLYTAKNLRR------SPVDITLIDKRNFHLFQPLLYQ
te  MSLKPGKASIYLQSPHHVVIVGGGFAGLEAAKQLGK------APVKVTLVDKRNFHLFQPLLYQ
cw1        MVAQSDSSNLHHVVIVGGGFGGLYAAQKLAK------APVKVTLIDKRNFHVFQPLLYQ
cw2         MDTKTLKNKTNVVIVGGGFGGLYTAQDLKN------SSVDVTLIDKRNFHLFQPLLYQ
su           VQPSLEKIHRVVIVGGGFGGLYAAKSLAR------APVTLTVIDRRNFHLFQPLLYQ
at  -APCASFGRQRHMQEHHVVVVGGGFGGLQLVHGLEG------APVRITLIDRRNHHLFQPLLYQ
(ec)           LTTPLKKIVIVGGGAGGLEMAT-QLGHKLGRKKKAKITLVDRNHSHLWKPLLHE
(ndc)-TAPRTYSWPDNKRPRVCILGGGFGGLYTALRLESLVWPEDKKPQVVLVDQSERFVFKPMLYE
```

Fig 3.5 (1/2) Alignment of Cyanobacterial sequences

N-terminus region including first nucleotide binding region
(top) Residue numbered according to *E. coli* residue.
Also shown: *E. coli* ('ec', Type I) and *Arabidopsis thaliana* **ndc** ('ndc')
Blue numbers distinguish multiple NDH2 genes within a species.

Species abbreviations (for accession numbers see Table 3.2):gv *Gloeobacter violaceus*, se *Synechococcus elongates* PCC 7942, ss *Synechocystis sp.* PCC 6803, ns *Nostoc sp.* PCC 7120, av *Anabaena variabilis*, pm *Prochlorococcus marinus*, mt *Mycobacterium tuberculosis*, ma *Mycobacterium avium*, ms *Mycobacterium smegmatis*, nf *Nocardia farcinica*, cg *Corynebacterium glutamicum*, rb *Rhodopirellula baltica*, cv *Calothrix viguieri*, te *Trichodesmium erythraeum*, cw *Crocosphaera watsonii*, su *Solibacter usitatus*, at *Agrobacterium tumefaciens*.

	nuc- binding site 2		region	
	173	185	327	339
gv1	..AALVGAGASGVEL..		..LGPSAQLAFQQAG..	
se1	..IAIVGAGPSGVEL..		..IPATAQAAFQQAN..	
ss1	..IAIVGGGYSGVEL..		..IPTTAQGAFQQTD..	
ns1	..VAIVGAGYSGVEL..		..VPATAQAAFQQAD..	
av1	..VAIVGAGYSGVEL..		..VPATAQAAFQQAD..	
se2	..IAVVGAGPSGIEM..		..QPGLAQVAYQQGA..	
pm	..LVIVGAGPTGVEL..		..WPPTAQVALQQGE..	
ss2	..VAIVGAGPAGVEM..		..KPALAQIAYQQGA..	
ns2	..VAIIGAGPSGIEL..		..QPPTAQVAYQQGI..	
av2	..VVVVGGGASGVEM..		..LPPTAQVAYQQGA..	
gv2	..VAVVGGGPAGIEL..		..QPTTAQAAYQEGE..	
mt1	..FVVVGAGPTGVEV..		..VPGVAQGAIQGAR..	
mt2	..FTVVGAGPTGVEM..		..VPGVAQGAIQGAK..	
ma1	..FVVVGAGPTGVEL..		..VPGMAQGAIQGAK..	
ma2	..FTVIGAGPTGVEM..		..VPGVAQGAIQGAK..	
ms	..FTVVGAGPTGVEM..		..VPGVAQGAIQGGR..	
nf1	..FVVIGAGPTGVEL..		..VPGQAQGAIQGAT..	
nf2	..FVVVGAGATGVEV..		..YPGQSPVAMQEGR..	
cg	..FVVVGAGPTGVEL..		..LPGVAQVAIQSGE..	
av6	..FTIVGGGATGVEM..		..LSGVAPEALQQGV..	
ns6	..FTIVGGGATGVEM..		..LSGVAPEALQQGV..	
rb	..FVIVGGGPTGCEL..		..VPGLAPAAMQMGT..	
av5	..FVIVGGGPTGVEL..		..LPGVAPVATQQGE..	
ns5	..FVIVGGGPTGVEL..		..LPGVAPVATQQGE..	
av3	..FVIVGGGPTGVEL..		..LPSVAPVAIQEGE..	
ns3	..FVIVGGGPTGVEL..		..LPSVAPVAIQEGE..	
cv	..FVIVGGGPTGVEL..		..LPGVAPVATQEGE..	
av4	..FVVVGGGPTGVEL..		..LPGVAPVAKQEGE..	
ns4	..FVVVGGGPTGVEL..		..LPGVAPVAKQEGE..	
se3	..FTIVGAGPTGVEL..		..LPGVAAVAMQQGA..	
ss3	..FVIVGAGPTGVEL..		..LPGVAPVAMQEAA..	
te	..FAIVGAGPTGVEL..		..IPGVAPAAMQEGF..	
cw1	..FVLVGGGPAGVEL..		..LPGIAPVAMQEGF..	
cw2	..FVIVGGGPTGVEL..		..LPGVAPVAMQQGA..	
su	..FVIVGGGPTGVEL..		..LPGVAPVAMQEGR..	
at	..FVIIGAGPTGVEM..		..VPGIAPAAKQQGA..	
(ec)	..IAIVGGGATGVEL..		..VPPRAQAAHQMAT..	
(ndc)	..VAVVGCGYAGVEL..		..LPTTAQVAFQEAD..	

Fig 3.5 (2/2) Alignment of Cyanobacterial sequences

Second nucleotide-binding site and "conserved region 3"

3.2.1.4 FMN-containing NDH2s

A 49kDa enzyme purified from *Acidianus ambivalens* mitochondria has exhibited NADH:quinone oxidoreductase activity but differs from all other NDH2 families in that it has a covalently-bound FMN in place of (non-covalently-bound) FAD prosthetic group of other NDH2 types. Homology searching revealed five additional *archea* species (table 3.5).

Sequence similarity to other NDH2s is minimal other than at the first nucleotide binding locus. (Sequence data is not presented here but may be examined in S.3.4)

Table 3.5 FMN-containing NDH2s

These enzymes, unlike all other NDH2s, have a covalently-bound FMN prosthetic group, and share little homology with the canonical NDH2.

FMN enzymes	
Acidianus ambivalens	AAM489504
Sulfolobus tokodai	BA000023
Sulfolobus solfataricus	AE006830
Picrophilus torridus	AE017261
Thermoplasma acidophilum	TACID4
Thermoplasma volcanium	BA000011

3.2.2 Fungal NDH2s

The genome sequence of *Saccharomyces cerevisiae* has been completed and three NDH2 enzymes (NDE1, NDE2 and NDI1) have been studied in this model organism. BLAST searching within fungal species identified putative NDH2s which are partitioned here into three groups. The data suggest that most fungal species encode three NDH2 genes. The most obvious difference from bacterial examples is an N-terminal extension of between 30 and 140 residues. Homology across fungal genes within this region is low suggesting that this is an intracellular targeting signal rather than an additional domain found in the mature protein. *Dictyostelium discoideum*, while not a fungus, also encodes three putative NDH2s that bear similarity to those of the yeasts, and has thus been included in this group (table 3.6, fig 3.6).

Type 2 fungal NDH2s incorporate, by definition, an EF-hand motif (similar to the plant **ndb** genes), while Type 1 retain the basic form of bacterial genes. Analysis of the Type 1 genes suggests two distinct groups, differing by short insertion/deletion regions and residue homology (see Supplementary Data S.4.2), so these are classified here as Types 1 and 1a. No functional distinctions can be inferred from these sequences differences; this probably reflects an evolutionary divergence.

Surprisingly, while *Saccharomyces cerevisiae* (like most fungi) expresses three NDH2 genes, all three of these are of Type 1. (*Neurospora crassa*, more typically, expresses one of each type.)

Table 3.6 Fungal NDH2s

	Type I	Type IA	Type II
Saccharomyces cerevisiae	**X61590 (*ndi1*)** Z47071 (*nde1*) Z74133 (*nde2*)		
Eremothecium gossypii	NP_985994 NP_984358		
Kluyveromyces lactis	XP_452480 AJ496546 (*nde1*)		CAH02955
Candida glabrata	XM_445076 CAG60223		
Candida albicans	CAL390487 XM_712959		
Neurospora crassa	XM_331371 (*nde2*)	EAA27430(*ndi*)	EAA32649 (*nde1*)
Gibberella zeae PH-1	XP_384306	XP_387439	EAA70552
Yarrowia lipolytica	YLAJ6852		CAG79173
Aspergillus nidulans	XP_405231	EAA62467	XP_411637
Aspergillus fumigatus	EAL90524	EAL90783	EAL87637
Magnaporthe grisea	XP_369761	EAA52307	EAA50381
Schizosaccharomyces pombe	Z99260 AL021837		
Debaryomyces hansenii	CR382136		
Ustilago maydis 521	XP_399779	EAK81746	EAK84847
Cryptococcus neoformans	XP_567114		AAW44492
Botryotinia fuckeliana			CAJ15142
Dictyostelium discoideum	AC116982 EAL72402		CAA98651

```
          5      nuc-binding site 1                                  63
Type 1
sci    52-KPNVLILGSGWGAISFLKHIDTKKYN---VSIISPRSYFLFTPLLPSAPVGTVDEKS
eg     56-KPNVVILGSGWGAISFLKHIDARKYN---VTVVSPRNYFLFTPLLPSTPVGTVDEKS
kl1    56-KPNVVILGSGWGAISFLQHIDAKKYN---VSIVSPRNYFLFTPLLPSTPVGTVDEKS
cg1    57-KPNVVILGSGWGAISFLKHIDTKKYN---VSIVSPRNYFLFTPLLPSAPVGTVDEKS
ca     93-KKTLVILGSGWGAISLLKNLDTTLYN---VVIVSPRNYFLFTPLLPSVPTGTVELRS
nce2  109-KKTLVVLGTGWGSVSLLKKLDTEHYN---VIVISPRNYFLFTPLLPSCTTGLIEHRS
gz1   101-KKTLVVLGSGWGSVGLLKNLDTENYN---VIVVSPRNYFLFTPLLPSCTTGLIEHRS
yl     99-KKTLVVLGSGWGSVSFLKKLDTSNYN---VIVVSPRNYFLFTPLLPSCPTGTIEHRS
an1    90-KKTLVILGTGWGSVSLLKKLDTENYN---VVVISPRNYFLFTPLLPSCTTGQVEHRS
af1    92-KKTLVILGTGWGSVSLLKKLDTENYN---VVVISPRNYFLFTPLLPSCTTGQVEHRS
mg1    96-RKTLVILGTGWGSVSLMKNLDVENYN---VIVISPRNYFLFTPLLPSCTTGTIEHRS
sp1    88-KKTLVVLGAGWGATSILRTIDTSLFN---VIVVSPRNYFLFTSLLPSTATGSVHTRS
sp2    87-KKNIVVLGSGWGAVAAIKNLDPSLYN---ITLVSPRDHFLFTPMLPSCTVGTLRLPS
sce2   91-KKELVILGTGWGAISLLKKLDTSLYN---VTVVSPRSFFLFTPLLPSTPVGTIEMKS
sce1  106-RKTLVILGSGWGSVSLLKNLDTTLYN---VVVVSPRNYFLFTPLLPSTPVGTIELKS
cg2   102-RKTLVILGSGWGSISLLKNLDTNIYN---VIVVSPRNYFLFTPLLPSTPVGRVELKS
ki2    94-RKTLVVLGTGWGSVSLLKHLDTSLYN---VIVVSPRNYFLFTPLLPSTPVGTVELKS
eg     80-RKTLVVLGSGWGSVTLLKNLDTTLYN---VVVVSPRNYFLFTPLLPSTPVGTVELKS
dh     86-KKTLVILGSGWGSISLLKNLDTTLYN---VVVVSPRNYFLFTPLLPSCPTGTVELRS
ca2   127-KKSIVILGSGWGAVSLLKNIDTSLYN---VSVVSPRNYFLFTPLLPSVPTGTVDMRS
um2   113-KKTIVVLGSGWGATSLLKNIDTEEYN---VVVISPHNYFLFTPLLPSVTVGTLDGRS
cn1   108-KPTLVVLGSGWGATSFLKTLDTDEFN---VVVVSPRNYFLFTPLLPSVTVGTLEPRS
dd1   107-RPKVVILGTGWGSLCFLRKLHTDLFD---VTIISPRNYFLFTPLLVGGTTGTVEVRS
dd2    42-NEKLIILGCGWGSYSFLKNLNSIKYD---ITVISPRNHFLFTPLLTSSAVGTLEFRS
Type 1a
gz2    29-ASTAALEGSGWAGYALAKKISPSAAS---CVLISPRSHFVFTPLIASTAVGTLEFRA
mg2    53-KERVVILGSGWAGYALARTLDPAKFD---RVVVSPRSHFVFTPLLASTAVGTLEFRA
nci1   42-KERVVILGSGWAGYSFAKDLDPEKYE---RIFISPRSYFVFTPLLASTAVGTLEFRT
af2    52-KERVVILGSGWGGYSLSRRLSPSKFA---PLIVSPRSYFVFTPLLTDAAGGSLDFSN
an2    34-KERVVILGSGWGGYTMSRKLSPKRFA---PVVVSPRSYFVFTPLLTDTAGGDLDFSH
um2    24-KQRLVVLGTGWGGYAFLKSLSYASLRRFDVKVISPTTSFSFTPLLAQASCATLDFRS
Type 2
an3    64-KPRLVILGTGWGSIALLKELNPGDYH---VTVVSPTNYFLFTPMLPSATVGTLGLRS
af3    64-KPRLVILGTGWGSIALLKNLNPGDYH---VTVVSPTNYFLFTPMLPSATVGTLGLRS
bf    164-KPKLVILGGGWGNVALLKTLNPDDYH---ITLVSPTNYFLFTPMLPSATVGTLEFRS
nce1  147-KPRLVILGGGWGSVALLKELNPDDYH---VTVVSPANYFLFTPMLPSATVGTLELKS
mg3   164-KPRLVILGGGWGGVAILKELNPEDWN---VTVISPANYFLFTPMLPSATVGTLELKS
gz3   162-KPRLVILGGGWGGVALLKELNPEDYH---VTVISPTNYFLFTPMLPSATVGTLELRS
kl3   162-KPKLVVLGSGWASVGLLKNLNPGDYD---VTVVSPQNYFLFTPLLPSAATGTLEVKS
yl2    92-KPRLVVLGSGWGSVALLNALKPGDYN---VTLISPSNYFLFTPMLPSATVGTLELRS
cn2   166-KPRLVIVGGGWGAVSLIQSLPAHAYN---VTLISPQTYFAFTPLLPSACVGTIEPRS
um3     ?-KERLVIVGGGWAAVGLLKSLDPEKYN---VTLISPNNYYLFNPLLPSAAVGTVEPRS
dd3    64-RERIIVLGTGWASLSFIQEIDLNKYE---IVVVSPRNYFLFTPMLTEATVGSVEVRS
```

Fig 3.6 (1/2) Alignment of fungal sequences

N-terminus region including first nucleotide binding region
(top) Residue numbered according to *E. coli* residue.

Blue numbers distinguish multiple NDH2 genes within a species. Those with idendified names are shown (eg. sce1 = *S. cerevisiae* nde1)

Species abbreviations (for accession numbers see Table 3.2):sc *Saccharomyces cerevisiae*, eg *Eremothecium gossypii*, kl *Kluyveromyces lactis*, cg *Candida glabrata*, ca *Candida albicans*, nc *Neurospora crassa*, gz *Gibberella zeae*, yl *Yarrowia lipolytica*, an *Aspergillus nidulans*, af *Aspergillus fumigatus*, mg *Magnaporthe grisea*, sp *Schizosaccharomyces pombe*, eg *Eremothecium gossypii*, dh *Debaryomyces hansenii*, um *Ustilago maydis*, cn *Cryptococcus neoformans*, dd *Dictyostelium discoideum*, bf *Botryotinia fuckeliana.*

```
                    nuc- binding site 2                region
          173            185        327               339
Type 1
sci1    .IVVVGGGPTGVEA..   ..LPPTAQVAHQEAE..  ..VFFDWIKLAFFKRDFFKGL*
eg      .IVVVGGGPTGVET..   ..LAPTAQVAHQEAE..  ..VIADWLKLTFFKRDCFKEM*
kl1     .IVVVGGGPTGVET..   ..LPPTAQVAHQQAE..  ..VIADWLKLAFFKRDCFKEL*
cg1     .IVVVGGGPTGVET..   ..LPPTAQVAHQQAE..  ..VVSDWLKLAFFRRDFFKDL*
ca      .IVVCGGGPTGVEA..   ..YPPTAQVAFQEGE..  ..VVLDWAKVYFFGRDCSKE*
nce2    .MVVVGGGPTGVEF..   ..YAPTAQVASQEGN..  ..VINDWVKSKLFGRDVSRE*
gz1     .MVVVGGGPTGVEF..   ..YAPTAQVASQEGS..  ..VAVDWLKSKAFGRDVSRE*
yl      .TVVVGGGPTGVEF..   ..YAPTAQVASQEGS..  ..VCIDWMKVRVFGRDISRE*
an1     .MVVVGGGPTGVEF..   ..YAPTAQVASQEGA..  ..VALDWVKAKLFGRDVSRE*
af1     .MVVVGGGPTGVEF..   ..YAPTAQVASQEGA..  ..VAADWLKAKIFGRDVSRE*
mg1     .VVVVGGGPTGVEF..   ..YGPTAQVASQEGA..  ..VINDWVKSKIFGRDVSRE*
sp1     .TVVVGGGPTGMEF..   ..YAPTAQVASQQGA..  ..VTLDWIRVKLFGRDISSL*
sp2     .ITVVGGGPTGMEF..   ..LPATAQVANQQGA..  ..VLIDWLKTRLFGRYDAKV*
sce2    .FVVVGGGPTGVEF..   ..FFPTAQVAHQEGE..  ..IAMDWTKVYFLGRDSSV*
sce1    .FVVVGGGPTGVEF..   ..LFPTAQVAHQEGE..  ..VAMDWAKVYFLGRDSSI*
cg2     .FVVVGGGPTGVEF..   ..LFPTAQVAHQEGE..  ..VAMDWTKAYFLGRDTSA*
kl2     .FVVVGGGPTGVEF..   ..LFPTAQVAHQEAE..  ..VAMDWAKVYFLGRDSSV*
eg      .FVVVGGGPTGVEF..   ..LFPTAQVAHQEGE..  ..VALDWLKVSFLGRDSSV*
dh      .IVVCGGGPTGVEV..   ..YAPTAQVAFQEGT..  ..VCLDWAKVSIFGRDCSKE*
ca2     .IVVCGGGPTGVEA..   ..YAPTAQVAFQQGI..  ..VCFDWIKVYLFGRDCSRE*
um2     .MVVVGGGPTGIEY..   ..FAPTAQAASQQGA..  ..VAADWFKVFLFGRDVSRE*
cn1     .MVVVGGGPTGVEY..   ..YAPTAQVASQQGI..  ..VLTDWLKVKLFGRDVSRE*
dd1     .FVVVGGGPSGVEF..   ..LAQTAQVASQQGR..  ..VSFDWLKSSVFGRDISRG*
dd2     .FVIVGGGATGIEF..   ..YPPTAQVASQSAV..  ..VPFDWMRTLIFVGIK*
Type 1a
gz2     .FAVVGGGPTGIEF..   ..LPATAQVASQQAT..  ..IPYYWLITWIFGRDISRF*
mg2     .FAVVGGGPTGIEF..   ..LPKTGQVASQQAV..  ..VPFFWFISWLFGRDISRF*
nci1    .FAIVGGGPTGIEY..   ..LPKTAQVAAQQAT..  ..VPIYWLVSWVFGRGISRF*
af2     .FAIVGAGPTGTEL..   ..PPATAQATFQEAK..  ..VPIYWLVSWVFGRGISRF*
an2     .FAIVGAGPTGTEL..   ..PPATAQVTAQEAK..  ..VAFRWLLNNLFGRDVSRY*
um2     .FVVVGGGPTGSEF..   ..LPATAQVASQQGT..  ..FFEPMCFMRIIMLADA-600
                                                ..VPANWASNLLFGRDVGRF*
Type 2
an3     .FVVCGGGPTGVEF..   ..LPATAQRANQQGQ..  ..LAMDWAKRALFGRVYCS-16
af3     .FVVCGGGPTGVEF..   ..LPATAQRANQQGQ..  ..LAMDWAKRALFGRGLFLR-5
bf      .FVVSGGGPTGVEF..   ..LPATAQRAHQQGQ..  ..LAMDWAKRTLFGRDLMNW*
nce1    .FVVCGGGPTGVEF..   ..LPATAQRAHQQGQ..  ..MAMDWGKRALFGRDLMSY*
mg3     .FVVSGGGPTGVEF..   ..LPATAQRAHQQGQ..  ..LAMDWLKRGLFGRDLMAY*
gz3     .FCVSGGGPTGVEF..   ..LPATAQRAHQQGQ..  ..MAMDWTKRGLFGRDLMSF*
kl3     .FVVCGGGPTGVEF..   ..LPATAQRAHQQGK..  ..LFMDWLKRGMFGRDILSE*
yl2     .FVVCGGGPTGVEM..   ..LPATAQRANQQGV..  ..MFQDWLKRGLFGRDFIAP*
cn2     .FVVCGGGPTGVEF..   ..YPATAQVASQQGK..  ..LMLDWVKRGIFGRDLSKF*
um3     .FVVCGGGPTGVET..   ..LPATAQVAAQQGH..  ..LLGDYIKRGIWGRDLSRI*
dd3     .FLVVGGGPTSIEG..   ..LPSTAQCASQQAK..  ..VSIDWVKTTLFGRDISRI*
```

Fig 3.6 (2/2) Alignment of fungal sequences

Second nucleotide-binding site, "conserved region 3" and the C terminus.

3.2.3 Higher plant NDH2s

These have been discussed above (3.1.2) and are tabulated here (table 3.7; S.5.1, S.5.2, S.5.3).

Table 3.7 Higher Plant and related NDH2s

Putative NDH2 genes in plant and homologous species (protozoan). **ndb** genes incorporate the EF-hand motif; **ndc** genes are most closely related to those of the cyanobacteria.

	nda	**ndb**	**ndc**
Arabidopsis thaliana	NM_100592 (*nda1*) NM_128553 (*nda2*)	NM_118962 (*ndb1*) NM_116741 (*ndb2*) NM_11826 (*ndb3*) NM_127645 (*ndb4*)	NM_120955 (*ndc*)
Oryza sativa	AK069755 AP003705	AK102878 AK100884 AK072710	AK064752
Solanum tuberosum	STU245861	STU245862	
Pisum sativum	AAO27256		
Lycopersicon esculentum	BT013064	BT013752	
Zea mays		AY107518 AY106876	
Brassica napus			CD837336
Leishmania major	AL512293		
Trypanosoma brucei	EAN78339		
Trypanosoma cruzi	EAN90346		
Plasmodium berghei		XP_673046	
Plasmodium falciparum		AL929356	
Plasmodium yoelii		EAA22988	
Theileria annulata		CAI75759	
Theileria parva		EAN31284	

3.2.4 Miscellaneous species NDH2s

Amongst other species there is little compelling evidence of NDH2 expression. However, there are protozoan species that encode a protein which most closely resembles those of the higher plant **ndb** group (*Plasmodium falciparum*: CAD51833, *Theileria annulata*: CAI75759 and related species), and the **nda** group, but with a 150 to 250 residue C-terminal extension (*Trypanosoma brucei*: EAN78339, *Trypanosoma cruzi*. EAN90346, *Leishmania major*: LMFP696).

No putative NDH2 sequences could be found in Genbank databases representing algal species, however a likely hypothetical protein is encoded in 'scaffold 117' of the (incomplete) *Chlamydomonas reinhardtii* genome[1] when using any of the four **ndb** protein sequences as queries (TBLASTN). Homology matches were reported in other scaffolds but visual inspection failed to identify a putative 'region 3' in these.

[1] http://genome.jgi-psf.org/chlre2/chlre2.home.html

```
          5        nuc-binding site 1                                  61
nda
at1   71-KPRVLVLGSGWAGCRVLKGIDTSIY------DVVCVSPRNHMVFTPLLASTCVGTLEFR
at2   69-KPRVVVLGSGWAGCRLMKGIDTNLY------DVVCVSPRNHMVFTPLLASTCVGTLEFR
os1   59-KARVVVLGTGWAGSRLMKDIDTTGY------EVVCVSPRNHMVFTPLLASTCVGTLEFR
os2  121-KPRVVVLGTGWAACRFLKDVDTRAY------DVVCISPRNHMVFTPLLASTCVGTLEFR
st    58-KPRIVVLGSGWAGCRLMKDIDTNIY------DVVCVSPRNHMVFTPLLASTCVGTLEFR
ps                                                          ?-TLEFR
le   107-KPRVVVLGSGWAACRFLKGIDTSMY------DVVCIAPRNHMVFTPLLASTCVGTLEFR
tb     8-KPKVVVVGTGWAGCYFVRDTKPQLA------ELHVLSTRNHHVLTPLLPQTTTGTLEFR
lm     9-KPNVVVLGTGWAGSYAAHHVDPNLC------NIHVISTRNHMVFTPLLPQTTTGTLEFR
tc     8-RPNVVVVGTGWAGAYFTRNLNCKLA------NLQVLSVRNHCVFTPLLPQTTTGTLEFR
ndb
at1   59-KKKVVVLGTGWAGISFLKDLDITSY------DVQVVSPQNYFAFTPLLPSVTCGTVEAR
at2   42-KKKVVLLGTGWAGTSFLKNLNNSQY------EVQIISPRNYFAFTPLLPSVTCGTVEAR
at3   42-KRKVVLLGTGWAGASFLKTLNNSSY------EVQVISPRNYFAFTPLLPSVTCGTVEAR
at4   63-KKKVVVLGSGWSGYSFLSYLNNPNY------DVQVVSPRNFFLFTPLLPSVTNGTVEAR
os1   62-KKKVVVLGTGWAGTSFLKDLDCSKY------EVKVISPRNYFAFTPLLPSVTCGTVEAR
os2   54-KKRVVIVGTGWAGASFLRNIDTSLY------DVHVVSPRNYFTFTPLLPSVTCGTVEAR
os3   55-KKKIVVLGTGWGGTTFLRNLDSRLY------DVQVISPRNYFAFTPLLPSVTCGTVEPR
tb    54-KKRVVVLGTGWGGTSFLKDVDISSY------DVQVVSPRNYFAFTPLLPSVTCGTVEAR
le    54-EKRVVVLGTGWGGTSFLKDLDISSY------DVKVVSPRNYFAFTPLLPSVTCGTVEAR
zm1                                                         ?-
zm2   54-KKKVVILGTGWAGASFLRNIDTSLY------DVHVVSPRNYFTFTPLLPSVTCGTVEAR
pf    39-KEKIIILGSGWGGFNFLLNIDFKKY------DVTLISPRNYFTFTPLLPCLCSGTLSVN
py    36-REKVVILGSGWGGIHFLLNIDFQKY------DVTLISPRNYFTFTPLLPCLCSGTLNVD
ta    42-KPKVLFLGSGWSSVFFIKNLNPKLF------DLTVISPRNYFTFTPLLPKILSGTVETN
tp    40-KPKVLFLGSGWSSVFFIKNLNPKLF------DLTVISPRNYFTFTPLLPKILSGMVESN
pb    35-REKVVILGSGWGGIHFLLNIDFQKY------DVTLISPRNYFTFTPLLPCLCSGTLNVD
(nc)111-KKTLVVLGTGWGSVSLLKKLDTEHY------NVIVISPRNYFLFTPLLPSCTTGLIEHR
ndc
at    80-RPRVCILGGGFGGLYTALRLESLVWPEDKKPQVVLVDQSERFVFKPMLYELLSGEVDVW
bn    82-RPRVCILGGGFGGLYTALRLESLVWPDDKKPQVVWVDQSERFVFKPMLCELLSGEVDVW
os    30-RPRVCILGGGFGGLYTALRLESLVWPNDKKPQVMLVDQSDRFVFKPMLYELLSGEVDVW
(ns)   4-TKRIVILGGGFGGLYTALRVSQLPWETQQKPEIVLVDQSDRFLFSPLLYELLTGELQSW
```

Fig 3.7 (1/2) Alignment of plant and related species sequences

N-terminus region including first nucleotide binding region
(top) Residue numbered according to *E. coli* homologous residue.

Blue numbers distinguish multiple NDH2 genes within a species. *Arabidopsis. thaliana* genes are identified according to recognised nomenclature(eg. at4 = **ndb4**).

Neurospora crassa (nc) also shown for **nda/ndb** comparison, and *Nostoc sp.* PPCC 7120 (ns) for **ndc** comparison.

Species abbreviations (for accession numbers see Table 3.4):at *Arabidopsis. thaliana*, os *Oryza sativa*, st *Solanum tuberosum*, ps *Pisum sativum*, le *Lycopersicon esculentum*, tb Trypanosoma brucei, lm *Leishmania major*, tc *Trypanosoma cruzi*, zm *Zea mays*, pf Plasmodium falciparum, py *Plasmodium yoelii*, ta *Theileria annulata*, tp *Theileria parva*, pb *Plasmodium berghei*, bn *Brassica napus*.

	nuc- binding site 2		region 3		C-terminus
	173	185	327	339	
nda					
at1	..CVVVGGGPTGVEF..		..LPALAQVAEREGK..		..VAINWLTTFVFGRDISRI*
at2	..CVVVGGGPTGVEF..		..LPALAQVAEREGK..		..VAINWFTTFVFGRDISRI*
os	..CVVVGGGPTGVEF..		..LPALAQVAERQGK..		..VAINWLTTLLFGRDISRI*
os	..CVVVGGGPTGVEF..		..LPALAQVAEREGR..		..VAVNWATTLVFGRDNTRIG*
st	..CVVVGGGPTGVEF..		..LPALAQVAERQGK..		..VLINWLTTLVFGRDISRI*
ps	..CVVVGGGPTGVEF..		..LPALAQVAERQGK..		..VAVNWGTTFVFGRDNSRIG*
le	..CVVIGGGPTGVEF..		..LPALAQVAERQGK..		..VAMNWGTTLIFGRDNTKIG*
tb	..TVVVGGGPTGVEF..		..LPTLAAVASRQGV..		..VIVNWLGSAIFGRDLTL-55
lm	..TVVVGGGPTGIEF..		..LPTLAAVASRQGR..		..VIVNWAGSQIFGRDITY-74
tc	..VVVVGGGPTGVEF..		..LPTLAAVASRQGA..		..VLVNWVGSAVFGRDTTF-89
ndb					
at1	..FVIVGGGPTGVEF..		..LPATAQVAAQQGA..		..VVSDWTRRYIFGRDSSRI*
at2	..FVVVGGGPTGVEF..		..LPATAQVAAQQGA..		..VVSDWMRRFIFGRDSSSI*
at3	..FVVVGGGPTGVEF..		..LPATGQVAAQQGT..		..VVSDWMRRFIFGRDSSRI*
at4	..FVVVGGGPTGVEF..		..LPATAQVASQQGK..		..VISDWTRRFVFGRDSSSI*
os	..FVIIGGGPTGVEF..		..IPATAQVAAQQGH..		..VVSDWTRRFIFGRDSSRI*
os	..FVVIGGGPTGVEF..		..LPATAQVASQEGA..		..VVSDWGRRFIYGRDSSSL*
os	..FVVVGGGPTGVEF..		..LPATAQVASQQGQ..		..VISDWSRRFIFGRDSSCI*
tb	..FVIVGGGPTGVEF..		..LPATAQVAAQQGT..		..VVGDWVRRYIFGRDSSRI*
le	..FVIVGGGPTGVEF..		..LPATAQVAAQQGT..		..VVGDWVRRYIFGRDSSRI*
zm	..FVVVGGGPTGVEF..		..LPATAQVASQQGQ..		..VVSDWTRRFIFGRDSSCI*
zm	..FVVIGGGPTGVEF..		..LPATAQVAAQEGS..		..VISDWGRRFIFGRDSSSI*
pf	..VAVVGGGPTGVEV..		..PTPTAQNAKQEAY..		..FFIDFIKTKWYGRPFIK*
py	..IVVVGGGPTGVEV..		..PPPTAQNAKQEAF..		..FIFSFLRTKIYGRPFI*
ta	..FLVVGGGPTGVEC..		..PFPTAQNAKQAAI..		..FFFDLLKNFFFSRHLILK-4
tp	..FLVVGGGPTGVES..		..PFPTAQNAKQEAI..		..FAFDMLKNFLFSRHLILK-3
pb	..VVIVGGGPTGVEV..		..?		
(nc)	..MVVVGGGPTGVEF..		..YAPTAQVASQEGN..		..VINDWVKSKLFGRDVSRE*
nd3					
at	..VAVVGCGYAGVEL..		..LPTTAQVAFQEAD..		..VRISCLAKSAVDSIALL-12
bn	..?				
os	..VAIVGLGYSGVEL..		..LPATAQVAFQQAD..		..VGISWFTKTAVDSLASL-12
(ns)	..VAIVGAGYSGVEL..		..VPATAQAAFQQAD..		..VGNFWLVRPIIETIYSA-12

Fig 3.7 (2/2) Alignment of plant and related species sequences

Second nucleotide-binding site, "conserved region 3" and C-terminus
?: sequence not determined

3.3 N-terminal region

In all eukaryotic species there is an N-terminal extension of around 50 to 200 residues when compared to prokaryotic examples. Homology between any pair of NDH2s in this region is poor, even for proteins which otherwise exhibit a high overall homology. Analysis of predicted secondary structure (Chapter 5) also indicates minimal conservation in this region. As eukaryotic NDH2s have been shown to be (at least predominantly) mitochondrially located proteins it must be concluded, at least in part, that this region constitutes a mitochondrial targeting signal. However any further analysis of this region is necessarily speculative.

Web-based servers for the prediction of subcellular targeting (MitoProt, PSORT) produced conflicting results for both intracellular location and mature protein N-terminal cleavage. In no case was an NDH2 predicted to have a mature N-terminus coincident with that of the prokaryotic enzymes, although empirically this might have been expected to be a possibility. The experimentally determined mature N-termini of the fungal NDI1 proteins of *S. cerevisiae* (de Vries *et al.*, 1992) and *N. crassa* (Duarte *et al.*, 2003) reveal cleavage of some but not all of the supposed target sequence, broadly consistent with the Mitoprot predictions.

Examining the proteins with known locations (ie. external- or internal-facing) it may be noted that the N-terminal region of the internal facing *S. cerevisiae* (NDI1) is shorter than those of the external counterparts (NDE1, NDE2); possibly the enzymes are initially imported to the matrix, and then re-exported by a second signal sequence. However the differences in leader length of genes with experimentally determined membrane destinations is not sufficient to provide a basis for reliable prediction.

3.4 EF-hand motif

The distinguishing feature of **ndb**-type enzymes is the insertion of around 70 residues towards to the C-terminus when compared to other NDH2s. BLASTP search of the NCBI CD (Conserved Domain) database using the four *Arabidopsis* **ndb** proteins identified an 'EF-hand' motif in this region for each of these.

The canonical EF-hand motif adopts a helix-loop-helix structure in which a single calcium ion becomes coordinated to conserved residues within the loop (Bhattacharya *et al.*, 2004; Luan *et al*, 2002). Conformational changes in the enzyme, resulting from calcium binding, provide a mechanism for regulating the kinetics of the enzyme.

Calcium signalling employing the EF-hand motif plays a significant role in *Arabidopsis* with as many as 250 putative examples identified (Day *et al.*, 2002).

EF-hands generally exist as contiguous pairs (or multiples of two). Alternative 'classes' of the motif however exhibit variation in both sequence and structure, and may bind alternative divalent cations such as magnesium (Nakayama *et al.*, 2000).

The canonical motif comprises 29 residues: 1-10 (helix'E'), 11-18 (loop) and 19-29 (helix'F').

By convention conserved calcium binding residues are termed X, Y, Z, -X, -Y, and -Z, and are located in or near the loop region (fig 3.8).

```
 1 2 3 4 5 6 7 8 9101112131415161718192021222324252627282 9
 H H H H H H H H H H L L L L L L L L H H H H H H H H H H H

                     X   Y   Z  -Y  -X    -Z
 E * * * * * * * * D * D * D G * I D * * E * * * * * * * *
                     N   N   N       N
                         S           S
```

Fig 3.8 Canonical EF-hand motif

H : helix; L : loop; X, Y, Z, -X, -Y, -Z : conserved residue loci names, characters below indicate residues (single letter amino acid notation), * represents any residue type.

Submission of each of the *Arabidopsis* **ndb** 80-residue sequences to MotifScan[1] for fold recognition revealed a putative pair of EF-hand motifs, though in each case only the first of each pair exhibits close homology to the canonical pattern. (Pattern matches for the second were identified only in the 'Prefile' databases, while the first was positively identified also in Pfam and Prosite). The alignment for the first hand of **ndb3** is shown below (fig 3.9).

```
 1 2 3 4 5 6 7 8 9101112131415161718192021222324252627282 9
 H H H H H H H H H H L L L L L L L L H H H H H H H H H H H H

                     X   Y   Z  -Y  -X    -Z
 E * * * * * * * * D * D * D G * I D * * E * * * * * * * * *
                     N   N   N       N
                         S           S

 D-I-A-A-I-F-K-K-A-D-K-E-N-S-G-T-L-T-M-K-E-F-H-E-V-M-S-D-I
```

Fig 3.9 ndb3 EF-hand

annotation as per fig 3.8.
lower line shows **ndb3** sequence in this region.

It is not possible to determine whether the second putative EF-hand contributes to calcium regulation in the **ndb** enzymes or if it a relic from an inherited domain maintained possibly for structural purposes. Interestingly, the nine residues linking the two putative hands (RYPQELYLK) are identical in all Arabidopsis **ndb**s.
Similar results were obtained using the putative EF-hand regions of fungal genes.

3.5 Intron/exon configuration

What benefits higher plants enjoy from expressing such a broad repertoire of NDH2s is not fully understood. *How* they acquired them can be partly answered by a survey of NDH2 expression in other species. The homology of **ndc** to cyanobacteria NDH2s is consistent with the hypothesis that this gene was acquired by an ancestral alga's symbiotic relationship with an endosimbiont cyanobacteria and the subsequent evolution of the latter into the chloroplast. Genes encoding chloroplast products are believed to have migrated to the nucleus and acquired chloroplast targeting signals. In the case of the **ndc** NDH2, a mitochondrial signal has been appended.

[1] http://myhits.isb-sib.ch/cgi-bin/motif_scan

An insight can be gleaned by inferring ancestral pathways through an examination the intron/exon configuration. Fig 3.2 has been annotated with the intron sites for *Arabidopsis* NDH2s. While the two **nda**s have seven identical intron sites, only three of the four **ndb**s are identical; **ndb4** is 'missing' four of the nine introns found in the other three. **ndc** also has nine introns but, with the exception of the site in 'Conserved Region 3', the loci are all different. Interestingly this intron site, uniquely, is shared by all *Arabidopsis* NDH2 genes. There is also EST evidence of an additional intron in the 3' UTRs.

Similar analyses were made of the other higher plant genes. One of the rice **nda**s has an intron pattern identical to that of the *Arabidopsis* **nda**s (although there is a 2-residue 'slippage' of the final site), but the other **nda** has 'lost' five of the introns, including that at the Conserved Region. The three **ndb**s share the nine *Arabidopsis* intron sites, though in two of these intron 1 is substantially extended (1478 nucleotides for one, compared to typically 390 for *Arabidopsis*). Introns of the **ndc** genes are identical.

As it is unlikely that both species have independently acquired introns at identical sites, a common ancestor probably expressed, at the very least, one **nda**, **ndb** and **ndc** gene, each punctuated by the full compliment of introns. It may be speculated further that the gene duplication events (twice in the case of **ndb**) occurred prior the acquisition of **ndc**, though a duplication/loss scenario is also possible. The 'loss' of introns from the *Arabidopsis* **ndb4** may be a chromosome-related phenomenon, as this is encoded in chromosome 2 while **ndb1**, **ndb2**, and **ndb3** are on chromosome 4. Similarly the two rice **nda**s are encoded on different chromosomes (1 and 7).

3.6 Summary of sequence analysis

Some 200 putative NDH2s have been identified here by TBLASTN searching, The collation of a substantial dataset has allowed a broad classification of these genes.

Sequence analysis failed to identify any significant homology between the FMN-containing group and other bacterial NDH2s (other then the presence of a nucleotide-binding region at the N-terminus). The covalently-bound FMN NDH2s are found in bacteria that inhabit extreme environments (pH, temperature, sulphur), but sequence analysis does *not* support the suggestion that NDH2s in this group have evolved from the canonical form as an adaptation to environmental conditions. Furthermore, some species express both the FMN and canonical enzymes (eg. *Sulfolobus solfataricus*). The group is not discussed further in this section.

Significant homology across the remaining NDH2s, from bacteria to higher plants, suggests a common evolutionary history. However, some are distinguished as having evolved the incorporation of an EF-hand motif immediately upstream of "Conserved Region 3", though there is no evidence of this among bacteria (fig 3.10). As would be expected, eukaryotic NDH2s have a N-terminal extension (relative to bacterial species) consistent with a mitochondrial targeting signal.

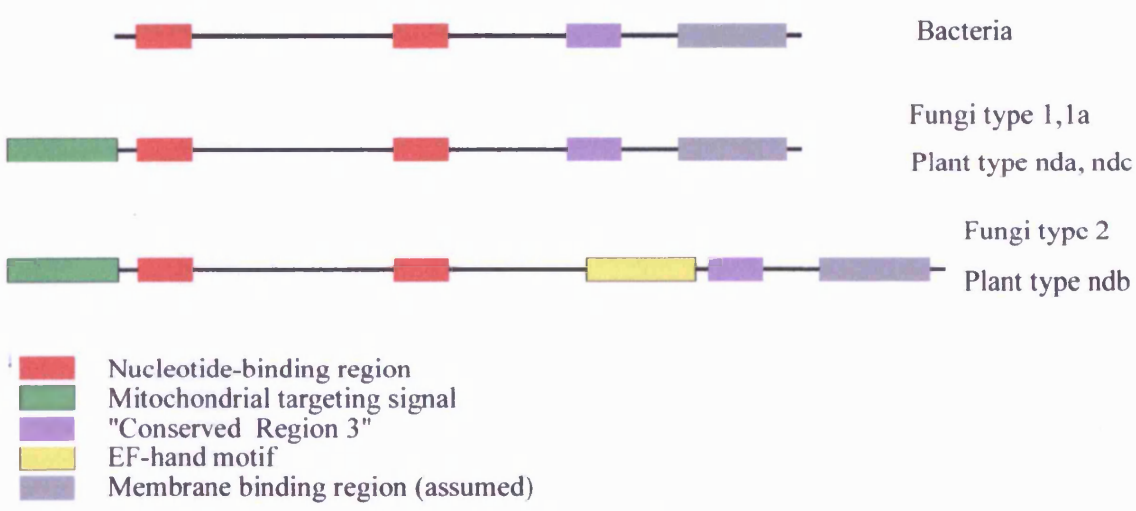


Fig 3.10 Schematic representation of NDH2 enzymes

This figure illustrates homologous NDH2 species (FMN-containing NDH2s are not included). The coloured loci are determined by similarity to known motifs.

The pattern that emerges from this survey is that the genomes of bacteria encode either one, two or three NDH2 genes (exceptionally six for some cyanobacteria, discussed further in Chapter 6) bound to the periplasmic membrane. Fungal species typically encode three genes, two bound to the external and one to the internal face of the mitochondrial inner membrane. One of these (an external-facing NDH2 in species examined experimentally), generally, is of the EF-hand containing motif type. In fact, *Saccharomyces cerevisiae* is untypical in that none of its NDH2s incorporates the EF-hand motif - a fact that has hitherto escaped observation. Higher plants encode as many as seven NDH2 genes, two of which incorporate the EF-hand motif. What benefits higher plants enjoy from expressing such a broad repertoire of NDH2s is a mystery.

There is however a small number of anomalous genes in this collation. Within the Type 2 bacteria, cyanobacteria and all trypanosome-type species there are genes with significant C-terminal extensions. These in turn fall into a larger subgroup of genes that depart slightly from the canonical sequence at Conserved Region 3 (see figures 3.4, 3.5 and 3.6). Also in this subgroup are genes from those cyanobacteria which have an 'excessive' NDH2 complement (mentioned above). It would be tempting to suggest that these are not indeed NDH2s, were it not for the fact that the gene of one of these (*Trypanosoma brucei*) has been cloned and characterised (Fang & Beattie, 2003). This anomaly is examined in detail in Chapter 6.

Chapter 4. Cloning and Expression of *Arabidopsis* **nda1**

4.1 Review of existing clones

The aim of this task was to obtain enzymatically functional *Arabidopsis* NDH2 by heterologous over-expression, for the ultimate purpose of obtaining crystals for X-ray structural determination.

Initial attempts to obtain functional NDH2 protein addressed four cDNA clones generously donated by Dr C. Moore (Flinders University, Adelaide, Australia). These clones (AT1, AT2, AT4 and AT6) contained inserts of putative *Arabidopsis* genes within the PCR cloning vector pGem (Promega), their orientation determined by partial sequencing. Previous research in this laboratory (Dr N. Fisher) had subcloned the AT6 insert into the expression vector pET22 (Invitrogen), initially in XL1 cells and subsequently transformed into BL21 host cells, thus producing an integrated system for the overexpression of **nda1**. The specifications of these strains are detailed in Table 4.1.

Table 4.1 Initial NDH2 strains

Columns (left to right) show clone name, *Arabidopsis* putative gene identifier, plasmid cloning vector type, orientation with respect to the MCS, and *E. coli* strain harbouring the plasmid.

clone	gene	NDH2	vector	orientation	host
AT1	2g29990	**nda2**	PGem-T	fwd	JM109
AT2	4g21490	**ndb3**	PGem-T	fwd	JM109
AT4	2g20800	**ndb4**	PGem-T	rev	JM109
AT6	1g07180	**nda1**	PGem-T	rev	JM109
AT611	1g07180	**nda1**	pet22b(+)	(fwd)	XL1
AT611	1g07180	**nda1**	pet22b(+)	(fwd)	BL21(DE3)

4.1.1 Plasmid Analysis

As both pGem and pET22 vectors incorporate ampicillin resistance genes these six strains were streaked onto LB-Amp plates and allowed to grow overnight. Two colonies from each of these were then picked and minicultures grown in 3ml LB-Amp media.

Plasmids preparations were obtained from the cultures using the Miniprep kit (Qiagen) and a preliminary analysis performed by restriction enzyme digestion and agarose gel electrophoresis. 4µl plasmid was digested with *Hinc*II (10U enzyme, NEB buffer 3, BSA, total reaction volume 20µl, 2 hour incubation at 37°), the fragments separated on a 1% agarose gel, and the latter photographed under UV illumination (fig 4.1).

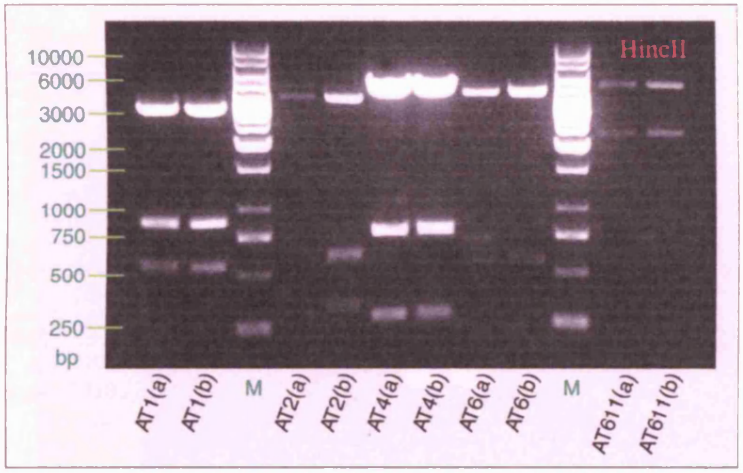


Fig 4.1. *Hinc*II digest analysis of AT clones.

Restriction digest of two colonies from each of the four original clones and AT611 (subcloned from AT6) showed unexpected fragment patterns for AT2 and AT4.

Predicted fragment sizes:

AT1 [3155, 835,554],
AT2 [3350, 470, 406, 252, 248],
AT4 [3982, 521, 263],
AT6 [4006, 544],
AT611 [4553, 2356].

M: marker lane; suffix in parenthesis denotes colony number.

Sequence files for the plasmids were assembled from published data for pGem-T and the appropriate NDH2 genes, and these were submitted to the NEB web-based digest predictor, NEBcutter[1], for comparison with gel results.

While AT1, AT6 and AT611 exhibited fragment patterns consistent with predicted data, those for AT2 and AT4 differed significantly from the expected results.
The digest was repeated with a different enzyme source (to eliminate the possibility of contamination) but the same gel pattern was observed. As the PCR primers used in creating the cDNA clones was not known there exists the possibility that additional *Hinc*II site(s) may be present in the plasmids, but this alone cannot explain the observed patterns.

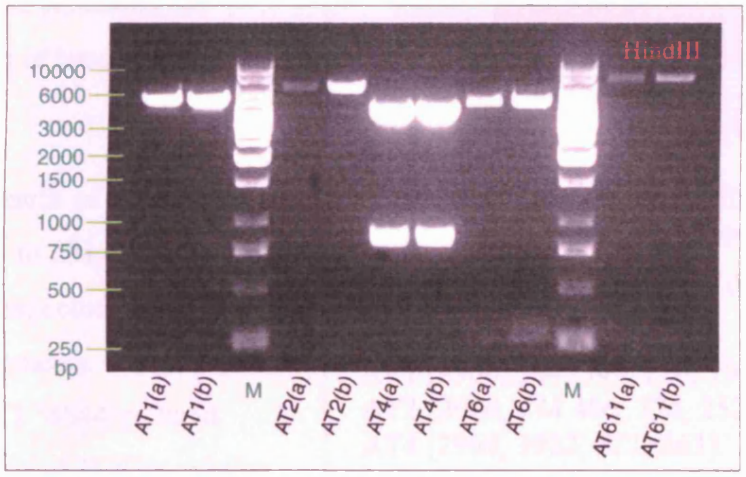


Fig 4.2. *Hind*III digest analysis of AT clones.

Restriction digest of two colonies from each of the four original clones and AT611 (subcloned from AT6) showed unexpected fragment patterns for AT2 and AT4.

Predicted fragment sizes:

AT1 [4544], AT2 [4730],
AT4 [not digested],
AT6 [4277, 273],
AT611 [6636, 273].

M: marker lane; suffix in parenthesis denotes colony number.

[1] http://tools.neb.com/NEBcutter2/index.php

A second digest using *Hind*III (10U enzyme, NEB buffer 2) was performed (fig 4.2) but the results were inconclusive. The single band of AT2 suggests either a single linearised fragment considerably larger than that expected (compare with the adjacent equivalent AT1 band), or an undigested plasmid, neither of which is consistent with predicted results. The two bands of AT4 could be explained as undigested supercoiled/nicked plasmid, though more generally three closer bands would be expected under these conditions. An identical gel was obtained in a repeat of this digest assay.

To resolve the uncertainty over AT4 a multiple digest was performed using *Hind*III, *Hinc*II and *Nco*I (NEB buffer 2, BSA) (fig 4.3). Multiple fragments were obtained though these were again inconsistent with predicted data while AT1, as a control, produced the expected pattern of bands.

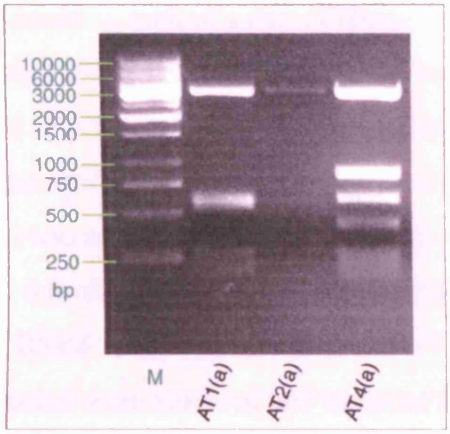


Fig 4.3 . *Hind*III-*Hinc*II-*Nco*I digest analysis of AT1, AT2, AT4 clones.

See text for interpretation of data.

Predicted fragment sizes:
AT1 [2960, 544, 544, 210, 195, 44, 27],
AT2 [2960, 474 406, 374, 252, 248, 17],
AT4 [2960, 1022, 521, 263].

An additional benefit of this multiple digest assay was to allay concerns over the higher than expected apparent mass of the larger fragments. In fig 4.1, for example, the AT1 '4544bp' band appears to be around 5500bp relative to the marker lane, suggesting that the plasmid may contain additional nucleotide pairs (possibly an immature mRNA which included introns). However, as the multiple digest assay produces smaller size fragments, and these exhibit reasonable correlation to the marker bands, this indicates that the size anomaly is a gel artefact.

It was concluded that the digest assays raised irresolvable doubts as to the integrity of AT2 and AT4, while AT1 and AT 6 (and AT611) may prove suitable candidates for protein expression experiments.

4.1.2 Preliminary expression of AT611

In order to determine whether the AT611(BL21) clone is a suitable expression system for the *Arabidopsis* 1g07180 gene, a preliminary expression assay was performed. FAD (the cofactor for NDH2) was added to the LB medium as endogenous *E. coli* host synthesis is likely to be inadequate for the anticipated expression level. The expression protocol was performed broadly in accordance with the manufacturer's recommended procedure (Promega).

From an overnight 5ml culture, 100µl was transferred to each of a pair of flasks containing 100 ml LB medium to which ampicillin (50 mg/ml) and FAD (5 µM) had been added, and these placed in a shaker at 37°C. At 20 minute intervals the cell density was determined spectrophotometrically at 600nm until the absorption reached 0.6 (A_{600} = 0.6), approximately four hours. IPTG (1 mM) was added to one of the flasks to induce expression and the cultures allowed to grow for a further three hours. Cells were then harvested by centrifugation and resuspended in 100ml buffer (TRIS 50 mM, NaCl 150mM, pH8.0). 10µl samples from each suspension were removed and analysed for protein content by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) (fig 4.4).

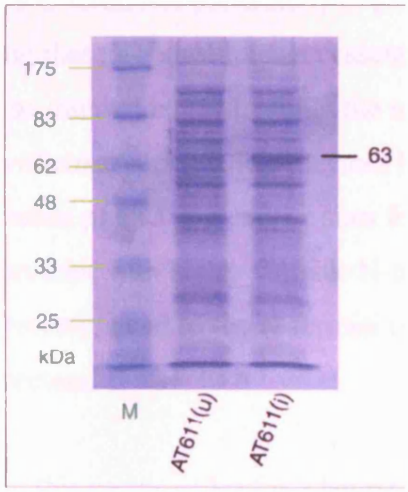


Fig 4.4. AT611 Protein Expression

SDS-PAGE analysis of Uninduced (u) and Induced expression (i) of AT611/BL21 clone demonstrating presence of a polypeptide of around 63 kDa.

M: protein marker NEB #P7708 (5µl).

A significant additional band was observed in the induced sample at around 63kDa, representing as much as 30% of the total protein content. This assay confirmed that pET22/BL21 is an efficient system for the synthesis of the NDH2 polypeptide, and provides confidence in AT611 as a correctly assembled plasmid.

4.2 Synthesis of a cDNA bookshelf

In order to compare the characteristics of different *Arabidopsis* NDH2 species it was decided that a cDNA 'bookshelf' a (cDNA library confined to specific genes) would be constructed due to the cost and limited availability of commercial clones. Genes for the seven NDH2s as well as the two homologous proteins (**ndd1** and **ndd2** - 5g22140 and 3g44190 respectively) were targeted for this procedure.

Seeds of *Arabidopsis thaliana* (strain col-0), kindly supplied by Dr A.Wingler (UCL), were grown to plants of around 10cm as source material for RNA extraction. The plants were then washed, and leaves and roots separately harvested and immediately frozen in liquid N_2. This material was stored at –80°C.

4.2.1 Selection of expression vector

pET32 was selected as a suitable expression vector for three reasons.

Firstly, analysis of primary structure of NDH2 genes from many species reveals a highly conserved C-terminus, particularly in position (length) and frequently in residue as well, suggesting that this constitutes an essential component of the higher-order structure, and possibly an internal location within the mature protein. By comparison, the mature N-terminus of eukaryotic NDH2 enzymes has not been determined experimentally for any species, while predicted cleavage sites for imported mitochondrial proteins suggests mature proteins with highly variable N-termini. Thus it appeared prudent to construct fusion proteins linked to the N-termini of the NDH2 proteins, an option possible with pET32 vectors.

Secondly, this vector codes for a large (109aa) soluble fusion protein (Trx), the primary purpose of which is to render a marginally insoluble protein soluble in the fusion product construction. NDH2 exhibits no putative transmembrane primary structure (hydrophilic α-helices for example) and is assumed to be a surface associated enzyme, and is thus potentially soluble in this vector construct.

Thirdly, in addition to the Trx domain, pET32 vectors incorporate both a six histidine polypeptide (for IMAC purification) and an enterokinase recognition site immediately upstream of the target gene insertion region. Thus it is also possible to cleave the mature target protein following successful folding in a soluble environment.
Finally, this fusion provides an alternative structure, quite distinct from to the C-terminus fusion of the AT611 clone.

4.2.2 Primer design for RT-PCR

Forward primers were designed to exclude the mitochondrial target signal peptides of the NDH2 genes, as predicted by MitoprotII. The first restriction site in the pET32 multiple cloning site (MCS), *Nco*I, was used in order to minimise additional N-terminal residues on NDH2 proteins following enterokinase cleavage, so suitable sites were incorporated near the primer 5' terminus, ensuring also that the reading frame is maintained across the resulting fusion peptide. Examination of the *Arabidopsis* NDH2 gene sequences reveals that some have endogenous *Nco*I sites, so for these *Bsp*HI or *Pci*I sites (which result in *Nco*I-compatible overhangs) were employed.

The endogenous NDH2 stop codon serves as the fusion product terminator so reverse primer design is a less demanding task. While *Nco*I (or compatible) sites could be incorporated also into the reverse primers, alternative MCS sites were employed (*Sal*I, *Hind*III, *Xho*I, *Sac*I or *Eco*RI), again ensuring no NDH2 endogenous site. This requires more restriction digest processing steps (of both vector and PCR product) but removes the possibility of reverse orientation insertion. (Vector/PCR ligation is the critical molecular engineering procedure.) In both cases standard PCR design considerations were also adhered to: compatible annealing temperatures (supplier data used), 3' C/G clamps, CG:AT ratio, avoidance of dimerisation artefacts (by visual inspection), extension beyond restriction site, and site choice selection to minimise incompatibility with gene sequence. The site and sequence data is summarised in table 4.2 and primer coding detailed in fig 4.5.

Table 4.2 Restriction sites for gene insertion into pET32(a)

abbrev: identifier used in PCR, primer and clone nomenclature
F: forward(5'), R: reverse(3') restriction sites.
sequence: Mitoprot II predicted native mature protein start (see Chapter 3)

ndh2	gene	abbrev.	site (F)	site (R)	sequence
nda1	1g07180	718	*Pci*I	*Sal*I	CTALQKQQVTDTVQ...
nda2	2g29990	999	*Pci*I	*Eco*RI	CTAQETQIQSPAKI...
ndb1	4g28220	822	*Pci*I	*Eco*RI	APLASKLLLLGTLS...
ndb2	4g05020	502	*Nco*I	*Eco*RI	KAFKDHPSLTRILV...
ndb3	4g21490	149	*Nco*I	*Xho*I	QAFHDYPSLSKILV...
ndb4	2g20800	080	*Nco*I	*Hind*III	SSLFKAYPSTSKIL...
ndc1	5g08740	874	*Bsp*HI	*Sac*I	VTNNSGTTEISDNE...
ndd1	5g22140	214	*Pci*I	*Hind*III	MEGIESGSKQGKRV...
ndd2	3g44190	419	*Nco*I	*Hind*III	MEKTESVSGKGKRV...

1g07180 (nda1)		**PCR length** 1461bp
718F	**CG** 0.55 **T_m** 70.8 **Site** *PciI*	
gene	..CTCTCGTCCCGTTTCTGCACCGCTCTGCAAAAGCAA..	
primer	TCGTCCCACATGTGCACCGCTCTGCAAAAGC	
PCR	TCGTCCAACATGTGCACCGCTCTGCAAAAGCAA..	
vector	..GACGACAAGGCCATGGCT..	
ligation	..GACGACAAGGCCATGTGCACCGCTCTGCAAAAGCAA..	
peptide	**Trx-H_6-S.tag-GTDDDDK|AMCTALQKQQ...**	
718R	**CG** 0.52 **T_m** 68.1 **Site** *SalI*	
gene	..CGGGACATTAGCCGAATCTGATCCATATCCCCTCCACTTCG	
Primer^C	GCCGAATCTGATCCATGTCGACTCCACTT	

2g29990 (nda2)		**PCR length** 1472bp
999F	**CG** 0.50 **T_m** 68.1 **Site** *PciI*	
gene	..CTCGCTTCTCGTTTCTGCACAGCTCAAGAGACTCAG..	
primer	CTTCTCACATGTGCACAGCTCAAGAGACTC	
PCR	CTTCTCACATGTGCACAGCTCAAGAGACTCAG..	
vector	..GACGACAAGGCCATGGCT..	
ligation	..GACGACAAGGCCATGTGCACAGCTCAAGAGACTCAG..	
peptide	**Trx-H_6-S.tag-GTDDDDK|AMCTAQETQ...**	
999R	**CG** 0.42 **T_m** 68.3 **Site** *EcoRI*	
gene	..TAAGGTTTTGTACAAAGTCTCAATCTCGGCCTACTGTAAATGGTGTCG	
Primer^C	CAAAGTCTCAATCTCGGCCTACTGTAAATGAATTCG	

4g28220 (ndb1)		**PCR length** 1711bp
822F	**CG** 0.42 **T_m** 69.5 **Site** *PciI*	
gene	..GCCTCTCGATCTGCTCCACTTGCCTCTAAGCTCCTCCTC..	
primer	CTCTCACATGTCTCCACTTGCCTCTAAGCTCC	
PCR	CTCTCACATGTCTCCACTTGCCTCTAAGCTCCTCCTC..	
vector	..GACGACAAGGCCATGGCT..	
ligation	..GACGACAAGGCCATGTGCCCACTTGCCTCTAAGCTCCTCCTC..	
peptide	**Trx-H_6-S.tag-GTDDDDK|AMSPLASK...**	
822R	**CG** 0.41 **T_m** 67.1 **Site** *EcoRI*	
gene	..AGGGATTCAAGCCGCATCTGACTCAATACTGCATTTTGTAATGTGT	
Primer^C	CCGCATCTGACTCAATACTGCATTTTGAATTCTG	

Fig 4.5 (1/3) *ndh2* primer design

GREY denotes nucleotides outside required coding region; RED: restriction sites;
GREEN: vector-encoded elements; Underline: delineates codon triplets
PCR: PCR product (5' terminus);
ligation: digested PCR product ligated to digested vector;
peptide: N-terminal elements in final expressed protein including
DDDDK| :enterokinase recognition and cleavage sites;
Primer^C: complementary representation of reverse (5') primer.

4g05020 (ndb2)				**PCR length** 1744bp
502F	**CG** 0.42	**T_m** 69.5	**Site** *NcoI*	
gene	. . GAGAGATTCTCTAAAGCTTTTAAAGATCATCCTTCTCTCACCAGG . .			
primer	GATTCTCCATGGCTTTTAAAGATCATCCTTCTCTCACC			
PCR	TCGTCCACCATGGCTTTTAAAGATCATCCTTCTCTCACC . .			
vector	. . GACGACAAGGCCATGGCT . .			
ligation	. . GACGACAAGGCCATGGCTTTTAAAGATCATCCTTCTCTCACC . .			
peptide	**Trx-H_6-S.tag-GTDDDDK|AMAFKDHP...**			
502R	**CG** 0.42	**T_m** 67.0	**Site** *EcoRI*	
gene	. . GGTAGAGATTCCAGTAGCATCTGATTTTCAACCTCCAAGGCTCCA			
Primer^C	CAGTAGCATCTGATTTTCAACCTCCAGAATTCC			

4g21490 (ndb3)				**PCR length** 1742bp
149F	**CG** 0.42	**T_m** 69.5	**Site** *NcoI*	
gene	. . GAAAGATTATCCCAAGCTTTCCATGATTATCCTTCTCTCTCC . .			
primer	GATTATCCATGGCTTTCCATGATTATCCTTCTCTCTCC			
PCR	GATTATCCATGGCTTTCCATGATTATCCTTCTCTCTCC . .			
vector	. . GACGACAAGGCCATGGCT . .			
ligation	. . GACGACAAGGCCATGGCTTTCCATGATTATCCTTCTCTCTCC . .			
peptide	**Trx-H_6-S.tag-GTDDDDK|AMAFHDYP...**			
149R	**CG** 0.42	**T_m** 67.0	**Site** *XhoI*	
gene	. . GATTCAAGTCGCATTTGAAAATCTCATGTCTCTCTTACTCTC			
Primer^C	CGCATTTGAAAATCTCATGTCTCGAGTACTCTC			

2g20800 (ndb4)				**PCR length** 1734bp
080F	**CG** 0.48	**T_m** 68.2	**Site** *NcoI*	
gene	. . TACCAGAGAGCTTCTAGCTTATTCAAGGCTTACCCTTCTACT . .			
primer	CCATGGCTAGCTTATTCAAGGCTTACCCTTC			
PCR	CCATGGCTAGCTTATTCAAGGCTTACCCTTCTACT . .			
vector	. . GACGACAAGGCCATGGCT . .			
ligation	. . GACGACAAGGCCATGGCTAGCTTATTCAAGGCTTACCCTTCTACT . .			
peptide	**Trx-H_6-S.tag-GTDDDDK|AMAASLFK...**			
080R	**CG** 0.47	**T_m** 68.2	**Site** *HindIII*	
gene	. . GGCCGTGACTCTAGCAGCATCTAAGCCATTCTTCTCCATCATTTT			
Primer^C	CCGTGACTCTAGCAGCATCTAAGCTTTTCTTC			

Fig 4.5 (2/3) *ndh2* primer design

GREY denotes nucleotides outside required coding region; RED: restriction sites; GREEN: vector-encoded elements; Underline: delineates codon triplets
PCR: PCR product (5' terminus);
ligation: digested PCR product ligated to digested vector;
peptide: N-terminal elements in final expressed protein including
DDDDK| :enterokinase recognition and cleavage sites;
Primer^C: complementary representation of reverse (5') primer.

5g08740 (ndc1)				**PCR length** 1418bp
874F	**CG** 0.50	**T_m** 70.5	**Site** *BspHI*	
gene	. . CTTTCGAGAGCAGTGACAAACAACAGTGGCACCACAGAG . .			
primer	GAGAGTCATGACAAACAACAGTGGCACCACAGAG			
PCR	GAGAGTCATGACAAACAACAGTGGCACCACAGAG . .			
vector	. . GACGACAAGGCCATGGCT . .			
ligation	. . GACGACAAGGCCATGACAAACAACAGTGGCACCACAGAG . .			
peptide	**Trx-H_6-S.tag-GTDDDDK|AMTNNSGTT...**			
874R	**CG** 0.45	**T_m** 68.2	**Site** *SacI*	
gene	. . AACCAAGGTTTTGTCTGGTTCTTGATCTCTGTTTGTTTTGCTCTGG			
Primer^C	CCAAGGTTTTGTCTGGTTCTTGAGCTCTGTTTG			

5g22140				**PCR length** 1124bp
214F	**CG** 0.41	**T_m** 68.4	**Site** *PciI*	
gene	. . CTTTGAGCTTGAATCTGTAAAAATGGAAGGCATTGAATCTGGGTC . .			
primer	GCTTGAATCTGTAAACATGTAAGGCATTGAATCTGGG			
PCR	GCTTGAATCTGTAAACATGTAAGGCATTGAATCTGGGTC . .			
vector	. . GACGACAAGGCCATGGCT . .			
ligation	. . GACGACAAGGCCATGTAAGGCATTGAATCTGGGTC . .			
peptide	**Trx-H_6-S.tag-GTDDDDK|AMMEGIESG...**			
214R	**CG** 0.39	**T_m** 67.2	**Site** *HindIII*	
gene	. . CAAGAGGCCTAAATCCTAAACTTGTATAAgtttgattcctattgtgtt			
Primer^C	GAGGCCTAAATCCTAAACTTGTATAAGCTTGATTCC			

3g44190				**PCR length** 1139bp
419F	**CG** 0.36	**T_m** 66.0	**Site** *NcoI*	
gene	. . ATTTTTGAATCACGAAGACATGGAAAAAACAGAATCTGTA . .			
primer	TTTTGAATCACGAAGCCATGGAAAAAACAGAATCTG			
PCR	TTTTGAATCACGAAGCCATGGAAAAAACAGAATCTGTA . .			
vector	. . GACGACAAGGCCATGGCT . .			
ligation	. . GACGACAAGGCCATGGAAAAAACAGAATCTGTA . .			
peptide	**Trx-H_6-S.tag-GTDDDDK|AMEKTESV...**			
419R	**CG** 0.44	**T_m** 68.3	**Site** *HindIII*	
gene	. . TGGATCCTAACCTTGTGGAGCCTTGAGTGATTGTGCTTATCTTCATGA			
Primer^C	CCTTGTGGAGCCTTGAGTGATTAAGCTTATCTTC			

Fig 4.5 (3/3) *ndh2* primer design

GREY denotes nucleotides outside required coding region; RED: restriction sites;
GREEN: vector-encoded elements; Underline: delineates codon triplets
PCR: PCR product (5' terminus);
ligation: digested PCR product ligated to digested vector;
peptide: N-terminal elements in final expressed protein including
DDDDK| :enterokinase recognition and cleavage sites;
Primer^C: complementary representation of reverse (5') primer.

4.2.3 Construction of pBS-718 and p32-718

Before attempting to obtain NDH2 clones from plant material plasmids were constructed to validate the PCR design and pET32 expression system using the AT6 clone as a template for the **nda1** gene.

nda1 was amplified from AT6 by PCR using 718F and 718R primers (1mM each). VENT polymerase was used with 2min extension times and an annealing temperature of 64°C. 5µl of the product was examined on an agarose gel which showed a concentrated band at the expected size, 1461bp (fig 4.6). This band was excised from the gel.

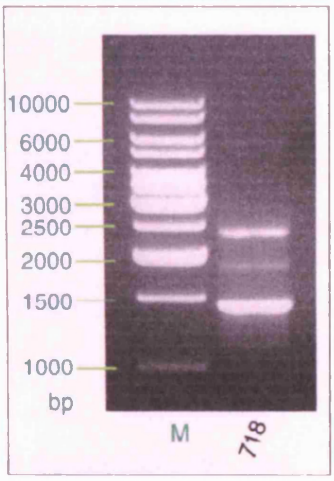


Fig 4.6. PCR amplification of 718 from AT6

A bright band is visible at the expected size for gene 1g07180 ('718'): 1461bp.

From an overnight miniculture of DH5α harbouring pBluescriptII, the plasmid was purified and 20µl subjected to a double digest using *Sal*I and *Eco*RV (NEB *Sal*I-buffer, BSA). 20µl of the PCR product ("718") was digested with *Sal*I (NEB *Sal*I-buffer, BSA), with separate *Sal*I and *Eco*RV digests of pBluescriptII as controls to confirm satisfactory activity of these enzymes, subsequently confirmed by gel analysis. Both pBluescriptII and 718 were purified from their respective digest products, their relative concentrations assessed by comparison on an agarose gel, and ligated using a 3:1 molar ration (insert:vector). (VENT PCR products have predominantly blunt ends, compatible with the *Eco*RV site in pBluescriptII.)

Competent DH5α cells were transformed with the ligation product and plated on LB-Blue medium. Six of the resulting white colonies were picked, cultured overnight and their plasmids extracted. (pBluescript without inserts produces blue colonies when grown in the presence of IPTG and X-gal.) One of these was subjected to multiple digest analysis (*Nco*I, *Nco*I/*Not*I, *Sal*I, *Pci*I, *Sal*I/*Pci*I) to confirm the integrity of this plasmid and the culture stored as a glycerol stock, "pBS-718". The 1461bp band from the *Sal*I/*Pci*I digest was excised from the gel and purified for subsequent ligation into pET32.

From an overnight miniculture of DH5α harbouring pET32(a), the plasmid was purified and 20µl subjected to a double digest using *Sal*I and *Not*I (NEB *Sal*I-buffer, BSA). The digest was purified then the terminal phosphate groups removed to minimise recircularisation of the plasmid, and samples of this and the digested 718 fragment examined on a gel to assess concentration (fig 4.7). 1µl insert was combined with 9µl vector (for an estimated 3:1 molar ratio) in a 20µl ligation for fifteen minutes.

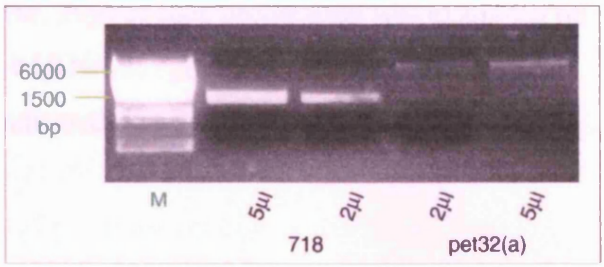


Fig 4.7. Ligation of pET32(a) and 718

Visual estimation of DNA concentration was used to determine 3:1molar ratio for ligation reaction. 2µl and 5µl samples of both linearised vector (pet32(a)) and PCR amplified gene (718) are shown.

Competent cells of BL21(DE3) Codon Plus (Novagen) prepared the previous day and stored at 4°C were transformed with the ligation product and colonies allowed to grow overnight on LB-Amp medium. Six colonies were selected, cultured overnight, and the plasmids isolated. Digest analysis confirmed correct incorporation of the **nda1** gene into pET32(a) and glycerol stocks made of this culture, "p32-718". Subsequent IPTG induction of this strain (section 4.3) confirmed that this construct synthesised a fusion protein of the predicted molecular mass.

4.2.4 Initial RNA extraction

RNA preparations were made using the Qiagen Plant RNeasy system. Three reactions were performed in parallel, two using leaf material and the other root, each using 100mg starting material. Each was eluted in a final volume of 60µl (RNase-free H_20). RNA concentration of the elutes was assessed spectrophotometrically by measuring the absorption at 260nm and using the approximation that an absorbance of 1.0 unit (A_{260}=1.0) corresponds to an

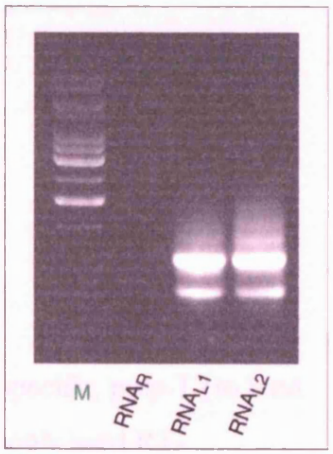


Fig 4.8. Gel analysis of RNA preparations.

The absence of 'streaking' in a 1% agarose gel suggests a 'clean' preparation

RNA concentration of 40µg/ml. Purity measurement (A_{260}/A_{280}) was not performed as this requires pH-buffered solutions. However, 20µl of each preparation was examined on an agarose gel for a qualitative assessment of RNA integrity (fig 4.8). The absence of 'streaking' below the bright rRNA bands indicated that the RNA had not been subjected to significant RNase degradation.

The two leaf preparation (RNA_{L1} and RNA_{L2}) were pooled. 20µl was taken from this and mixed with 980µl ddH_20 in a quartz cuvette (X50 dilution). The remaining 60µl was snap-frozen in liquid nitrogen and stored at –80°C. An absorbance value of 0.100 was measured for the leaf sample.

A similar test of RNA_R, the root sample, yielded and absorbance <0.005.It had not been possible to wash all the compost from the knotted root network and this contamination may explain the failure to extract RNA, although experimental error cannot be eliminated. However it was decided to defer root RNA preparation, and to consider using root material directly from agar-plated seeds at a later date.

The leaf RNA sample elute was thus determined to be 200µg/ml ("R1") so the preparation had isolated 12µg RNA from each 100mg leaf material, of which 12µg remained in 60µl. This recovery compares reasonably with an optimum 35µg from 100mg *Arabidopsis* leaves, claimed in the Plant RNeasy documentation.

4.2.5 RT-PCR

Reverse Transcription reactions used three primer types: gene-specific, poly-T (to bind eukaryotic mRNA tails), and random primer mixes. Two commonly used RT-polymerses are Omniscript and Sensiscript, the latter appropriate for low concentration template, both available from Qiagen (W.Sussex, UK).

Gene-specific primers and a T_{16} primer (a 16 'T' oligonucleotide) were obtained and dissolved to 100μM in ddH_20. A mixture of the nine reverse sense primers at 10μM each was made ("R-primers"), and a similar mixture for the forward primers ("F-primers"). The initial 20μl RT reaction used Omniscript enzyme with the poly-T primer (1μM) and 1ug RNA and the product named "RT-1". Initial PCR was attempted using the proofreading VENT polymerase with parallel reactions containing 2mM and 5mM Mg^{++} respectively. 2μl RT-1 was used as template with 1 μl each of F-primers and R-primers with 2 minute elongation times and an annealing temperature of 64°C. As controls 1ug pBS-718 was used as a template, also at both Mg^{++} concentrations. PCR product was obtained from the control (5mM Mg^{++}), but from no other reaction. PCR reactions were repeated under a variety of altered conditions: lower annealing temperature (54°C), 718F/718R primers alone, altered Mg^{++} concentration, increased cycling (45 cycles), substituting Taq polymerase for VENT, but without any bands of the predicted sizes being observed on the gels.

PCR product of the correct size was finally obtained when Sensiscript was substituted for Omniscript at the Reverse Transcription stage. In the successful procedure RNA was diluted eight-fold in DNase-free water and denatured (65°C for 5 minutes) prior to the addition of 2μl (50ng RNA) to the reaction tube. (This is the maximum recommended by the manufacturer.)

Four RT reactions were performed in parallel using different primers: T_{16}, X (random primers) and two gene-specific primers, 718R and 822R each at a final concentration of 1μM. Four PCR reactions were attempted using these RT-products as templates (table 4.3). In each case Taq (4U) was selected as the polymerase with primers at a concentration of 1μM each, and 3μl template in a 50μl reaction volume. Forty cycles were performed with an annealing temperature of 60°C.

Table 4.3 Primers used in 718 and 822 PCR reactions

cDNA: name assigned to RT-PCR product; RT template: RT product used in PCR reaction;

cDNA	RT template	PCR primers
718_1	718R	718F, 718R
718_2	T_{16}	718F, 718R
718_3	X	718F, 718R
822	822R	822F, 822R

While no product was detected for 822, gel analysis of 10µl samples (fig 4.9) revealed potentially valid products for each of the 718 reactions. Perhaps surprisingly, the most prominent band was observed where random primers had been used in the RT stage (718_3 in fig 4.9). This band ("P718") was excised from the gel, purified and stored for later cloning.

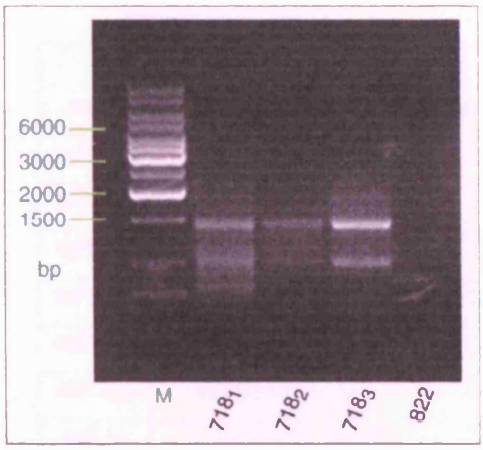


Fig 4.9. RT-PCR of 718 and 822

Bands are visible at around 1461bp (the expected size) for all three 718 RT-PCR reactions but not for the 822 RT-PCR.

PCR reactions were then attempted using the random primer derived template with seven primer pairs under the same reaction conditions. On this occasion (fig 4.10) a significant product was observed for 718 but, at around 1kb, this was rejected. However, weak bands of the approximately correct mass can be observed for 999 (excised, "P999") and 874. Further PCR attempts obtained similar weak bands for 502, 214, 419 (data not shown).

The inability to obtain PCR products of the correct size, even with low annealing temperatures, suggests failure of the Reverse Transcription stage under the above conditions. This may be a consequence of the low expression level of NDH2 mRNA. Since mRNA represents only around 3-5% total RNA the RT reaction may be improved by prior isolation of mRNA from total RNA.

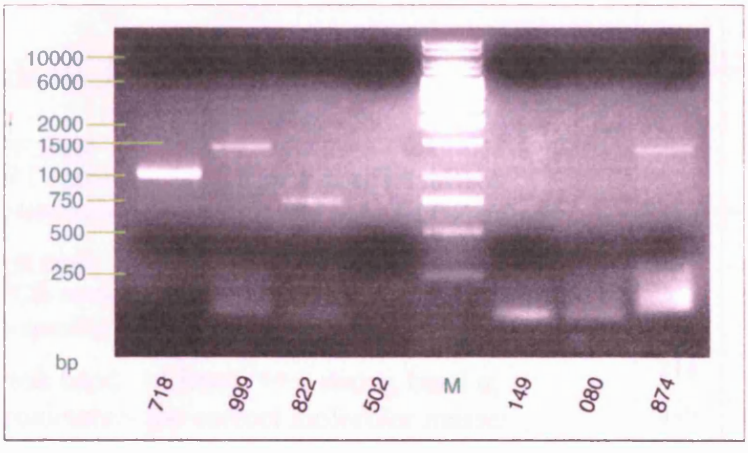


Fig 4.10. RT-PCR of 718, 999, 822, 502, 149, 080 and 874

A significant band is visible for 718 but at too low a mass in this reaction (approx 1kb). Potentially valid bands are however obtained for both 999 and 874 (approx 1.5kb).

Two further larger-scale RNA preparations were made (as described above), each from 300mg leaves (three RNeasy reactions each) yielding 200µl preparations ("R3" and R4") with RNA concentration assessed as 278µg/ml and 210µg/ml respectively.
mRNA was isolated using the Qiagen Oligotex kit yielding two 50µl mRNA preparations ("mR1", "mR2" respectively), each estimated to contain around 2µg mRNA.

Ten Reverse Transcription reactions were then performed, each using 200ng mRNA template, as detailed in table 4.4. A 1:1 mixture of Omniscript and Sensiscript was used in these reactions, otherwise reagents were as described above. PCR reactions were then performed using combinations of primer pairs and RT template, and gel analysis revealed putative valid products as indicated in the table.

RT reactions

name	primer	mRNA	quantity
$mR1_{999}$	999R	mR1	1
$mR1_{080}$	080R	mR1	1
$mR1_{149}$	149R	mR1	1
$mR2_{874}$	874R	mR2	1
$mR2_{502}$	502R	mR2	1
$mR2_{822}$	822R	mR2	1
$mR1_T$	T_{16}	mR1	2
$mR2_T$	T_{16}	mR2	2

PCR reactions

gene	primers	template	result
999	999F/999R	$mR1_{999}$	++
		$mR2_T$	
		$mR1_{999}$	
080	080F/080R	$mR1_{080}$	
		$mR2_T$	
		$mR1_{080}$	
		$mR1_T$	
		$mR2_T$	
149	149F/149R	$mR1_{149}$	
		$mR2_T$	
		$mR1_{149}$	
		$mR1_T$	
		$mR2_T$	
874	874F/874R	$mR2_{874}$	
		$mR1_T$	+++
		$mR2_{874}$	++
502	502F/502R	$mR2_{502}$	
		$mR1_T$	
		$mR2_{502}$	+
822	822F/822R	$mR2_{822}$	++
		$mR1_T$	
		$mR2_{822}$	+
214	822F/822R	$mR2_{822}$	+++
419	419F/419R	$mR2_{419}$	+++

Table 4.4 RT-PCR reaction table

Upper pane
ten RT reations using Oligotex-purified mRNA template with gene-specific and poly-T (T_{16}) primers.

Lower pane:
24 PCR reations using the above RT products with gene-specific primer pairs.

+ weak band, ++ band, +++ strong band at approximately the correct molecular masses.

4.2.6 cDNA cloning

P718 was digested with *Pci*I and *Sal*I (NEB buffer3, BSA) and the product purified. Similarly P99 was digested with *Pci*I and *Eco*RI (NEB buffer 3). Two10µl aliquots of pET32(a) were digested with the same enzyme pairs and then phosphatased to prevent recircularisation. All four samples were heated to 65°C for 20 minutes to inactivate the enzymes.

Following analysis on a gel to determine molar concentration, PCR products were ligated to their respective plasmids (3:1 insert:plasmid) and transformed into DH5α cells. Cultures were grown from selected colonies and the purified plasmids examined by digest analysis using *Hinc*II and *Pvu*I. The fragment pattern for the P999 insertion was not as expected for any of the four colonies examined, but two of the four p718 patterns were consistent with predicted digest sizes. One of these was stored ("p32-718A").

Five of the PCR products derived from the mRNA template were selected for cloning into pBluescript: 822, 874, 502, 214 and 419. As all were of relatively low concentration they were first amplified by VENT PCR and then ligated as described above. Digest analysis of the resulting plasmids suggested that one of the 874 colonies was correctly constructed (using *Bsp*HI), though none of the colonies selected from the other constructs could be so verified.

4.3 Expression of AT611 and p32-718

4.3.1 Introduction

The primers and construction of p32-718 have been described above. The primers for AT611 (Dr N Fisher, this laboratory) are:

7180A AAAAGGATCCGATGTTGTGGATCAAAAACCTAG restriction site: *Bam*HI

7180B AATTAAAAGCGGCCGCGATTCGGCTAATGTCCCGAC restriction site: *Not*I

and the digested PRC product inserted at *BamHI/NotI* sites in pET22(b).

The AT611 fusion protein has an N-terminus **pelB** *E. coli* signal which targets the protein to the (more oxidising) periplasm, where disulphide bonds are more likely to form. Proteins thus exported have their signal sequence removed at the "perplasmic peptidase site" (fig. 4.11). A six-histidine sequence at the C-terminus (H_6) is appended for IMAC purification. The full-length **nda1** protein (including putative mitochondrial target sequence) is encoded.

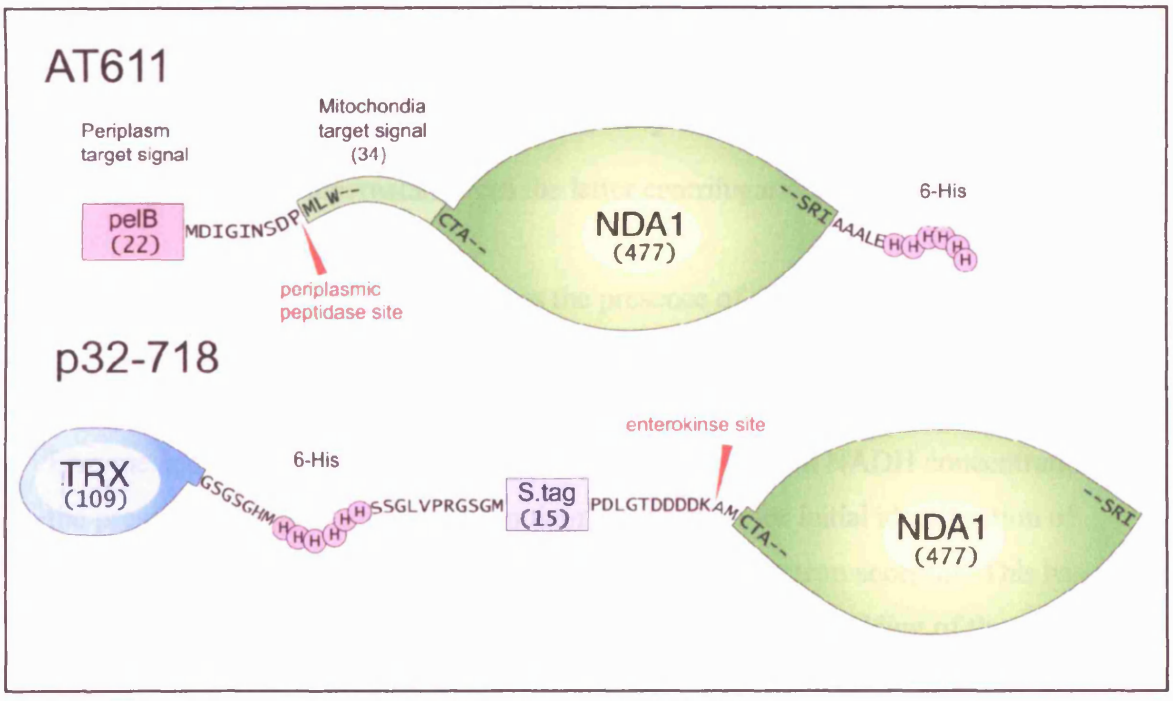


fig 4.11 AT611 and p32-718 expressed fusion proteins

Symbolic representation of fusion products for each construct showing positions of His-tags the cleavable pelB signal (AT611), soluble Trx domain, S-tag and enterokinase site (p32-718). Figures in brackets show number of amino acid residues.

The soluble TRX domain encoded in pET32(a) is intended to promote solubility in marginally soluble target proteins, while the enterokinase site provides the option of cleaving the target protein (for example while bound to an IMAC column). When used with a host strain encoding *trxB* (eg. BL21*trxB*, Invitrogen) the presence of cytoplasmic thioredoxins may also promote disulphide bond formation; however the host strain used here (BL21 Codon Plus(DE3)–RIL, Invitrogen) does not encode *trxB*. The forward PCR primer was designed so as to exclude the putative mitochondrial target signal from the **ndal** protein domain (Mitoprot II), and the reverse primer to retain the native **ndal** C-terminus in the fusion product. Predicted masses for the expressed proteins[1] were 61kDa for AT611 (58kDa after **pelB** cleavage) and 70kDa for p32-718.

In expression experiments *E. coli* cells were harvested (10,000g, 15min, 4°C)[2], resuspended in 30ml 20mM KPi buffer pH8.0 then ruptured by the French Press technique. The buffer contained no protease inhibitors as the effects of these on **ndal** activity is unknown. This was then separated into three fractions by centrifugation:

> Inclusion Bodies (3,000g, 15 min, 4°C, resuspended in 30ml KPi by homogenisation). This contains aggregated protein and unruptured cells.
>
> Membrane (120,000g, 60min, 4°C, resuspended in 30ml KPi).
>
> Soluble (the supernatant from the latter centrifugation).

SDS-PAGE analysis was used to assess the presence of expressed protein in the fractions.

Enzyme activity can be quantified by measuring the change in NADH concentration in the presence of a (synthetic) soluble quinone. However, for initial identification of activity potassium ferricyanide (FeCN) was used as the electron acceptor. This has the advantage, in addition to cost, of requiring only partial correct folding of the enzyme (the FAD and NADH pocket domain(s)).

---

[1] www.expasy.org/peptidetmass

[2] all centrifugation forces are average g units (g_{AV}) throughout this chapter.

The activity was determined spectrophotometrically. A change in absorbance rate of 0.01 units per second was determined to be sufficiently distinct from the background depletion rate of NADH in the presence of FeCN. As NADH exhibits a molar absorbtion coefficient of. 6.2×10^3 $M^{-1}cm^{-1}$ at 340nm this requires an NADH $\rightarrow$ NAD^+ reaction rate of 1.5μM s^{-1} [1.5nmol s^{-1} in a 1ml cuvette]. Assuming a typical enzyme electron transfer turnover of $50s^{-1}$ (NADH: $25s^{-1}$ as two electrons are removed per NADH molecule) this requires the addition of 60pmol of fully active enzyme. However, as the initial purpose was to detect trace levels of activity, minimum enzyme additions were generally two orders of magnitude greater than this – typically 50μl of a 100μM enzyme preparation (6mg/ml @ m_r = 60kDa).

4.3.2 Expression of active protein

As NDH2s are assumed to be weakly bound to membranes and lack putative transmembrane helices it was hoped that expression of one or other of the AT611 and p32-718 fusion proteins (as opposed to expression of **nda1** alone) would produce active, and possibly soluble, enzyme. The incorporation of a soluble fusion product in p32-718 was selected specifically for this purpose. It is possible that only a proportion of the synthesised protein would fold correctly (the majority aggregating as inclusion bodies). Consequently the products were assayed by activity rather than solely SDS-PAGE analysis of the fractionated protein.

100ml cultures of each construct were grown in 500ml flasks by inoculating BL medium including 5μM FAD with 200μl from overnight AT611 and p32-718 minicultures. (FAD is the essential NDH2 cofactor and it may be that supply of endogenous *E. coli* FAD does not satisfy the requirements of the overexpressed enzyme.) In accordance with the manufacturer's suggestions the cultures were allowed to grow to a density of OD_{600} = 0.6 at 37°C by taking samples every 20 minutes. Cultures were then each split into two 50ml samples, one of which was induced by addition of IPTG to 1mM. Culture growth was allowed to continue for a further three hours, then the final culture densities recorded (OD_{600}), the cells harvested, ruptured and fractionated as described above.

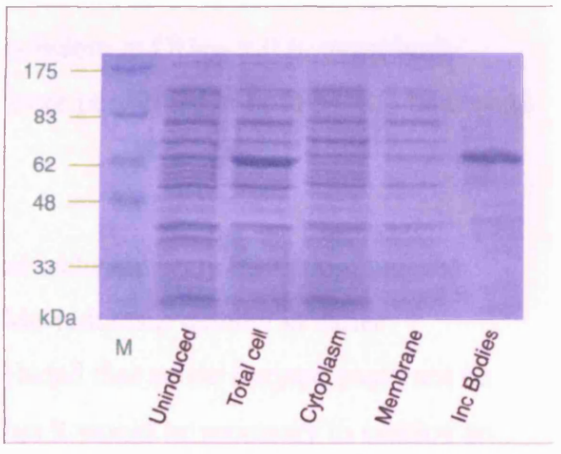


Fig 4.12 AT611 Protein Fractionation

High expression level is observed in the Inclusion Body fraction

Gel analysis of the AT611 fractions revealed that the protein was expressed predominantly or exclusively as inclusion bodies (fig 4.12), with similar results for p32-718. This conclusion was verified by the successful solubilisation of the fraction in 6M Guanidine and failure to solubilise it in 1% Tween-80 (ie. it was not a heavy membrane component).

However it is possible that, while predominantly expressed as inclusion bodies, some active enzyme may be present in the cytoplasmic or membrane fractions. Enzyme activity assays were performed on these fractions and compared against activity rates for similarly fractionated non-induced samples, the sample quantities (nominally 20µl) adjusted to compensate for differences in the recorded final cell density measurements. No significant differences were detected between induced and uninduced samples, indicating that no active enzyme was present.

Marginally soluble proteins can sometimes be produced in active form by varying the growth/induction protocol. These experiments were repeated a number of times using different FAD concentrations (zero, 0.1µM, 1µM, 10µM), growth temperature (20°C) and IPTG concentration (0.2mM, 0.5mM), and combinations of these.

An alternative induction protocol was attempted involving subjecting the cells to 'heat shock' prior to induction: upon reaching a density of OD_{600} = 0.25, glycerol (0.1%) and potassium glutamate (0.1mM) were added to the cultures which were then grown at 42°C for 30 minutes. Expression was induced as before at OD_{600} = 0.6, using both 0.25mM and 1mM IPTG. Final growth phases were performed at both 37°C/3 hours and 20°C/overnight.

In each case gel analysis of the fractions revealed inclusion body formation, and the enzyme assay failed to detect NADH-FeCN oxido-reductase activity in either cytoplasmic or membrane fractions. It was concluded that active enzyme could not be expressed directly using these constructs, and that it would be necessary to employ an alternative technique: refolding of the inclusion bodies.

4.3.3 Growth conditions and purification of inclusion bodies

Refolding of solubilised inclusion bodies requires a substantial supply of expressed protein. Following some disappointing yields in initial 500ml culture preparations, a brief examination was undertaken to asses the effects of anaerobic conditions and extended growth (to stationary phase) in minicultures. Anaerobic cultures were grown in sealed 10ml Sterilin® tubes filled to the lid; aerobic cultures were 5ml volume with loosely-fitting lids. The standard induction protocol was followed (1mM IPTG at OD_{600}=0.6) with no supplementary FAD in the LB medium. The yield was assessed by examining SDS-PAGE gels for AT611, p32-718 and pET32 (the native plasmid without gene inserted) with lanes loaded to provide similar intensity bands (fig 4.13).

This demonstrated that both fusion body (p32-718) and soluble peptide (pET32) yield is substantially compromised by extended growth times, whether due to protease activity or inefficient expression. In 'log phase' samples yield was reduced under anaerobic conditions as expected. There was also evidence of expression in uninduced samples when grown to stationary phase. Similar results were obtained for AT611 (data not shown).

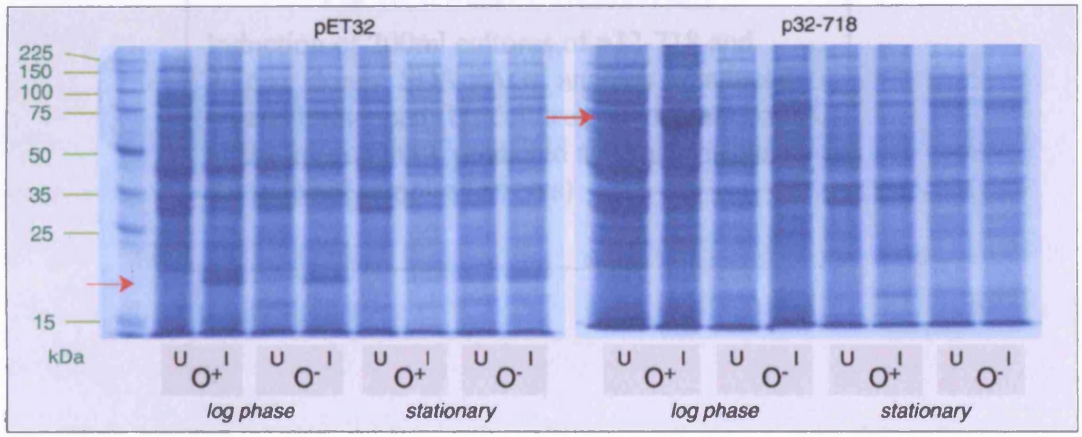


Fig 4.13 Effects of growth conditions

Expressed protein yield in pET32 and p32-718 systems. **U**: uninduced, **I**: induced, **O**$^+$: aerobic, **O**$^-$: anaerobic, ***log phase***: expression halted after 3 hours, ***stationary***: growth continued for 12 hours. (Gel loading adjusted to provide similar intensities.)
Red arrows identify predicted mass of expressed peptide.

Consequently all large scale cultures (200ml) were grown with loosely-fitting cotton wool plugs in 2 litre conical flasks and allowed to grow for 3 hours following induction. Gel analysis of these cultures (fig 4.14) confirmed high expression levels – as much as 50% of total protein for p32-718.

Inclusion bodies were purified as described above and suspended in 30ml 6M guanidine or 8M urea as required. Protein concentrations were determined by Bradford assay to be approximately 4mg/ml for p32-718 and 2.5 mg/ml for AT611.

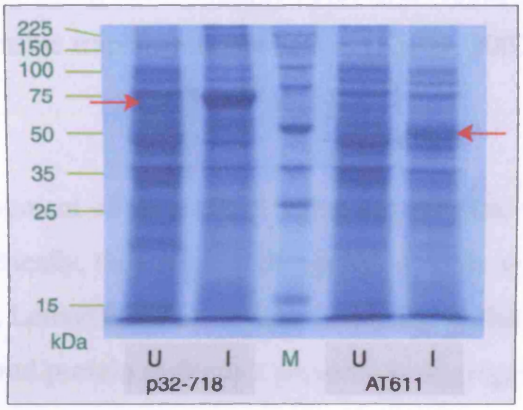


Fig 4.14. Large scale induction

Induction of 200ml cultures of p32-718 and AT611 clones. SDS-PAGE analysis of induced (I) and uninduced (U) cultures. Significant bands of the approximate predicted sizes are visible in the induced samples (arrows)

4.4 Refolding of inclusion bodies

Many exogenously expressed proteins aggregate as incorrectly folded inclusion bodies when over-expressed *E. coli*, generally within the cytoplasm but also in the periplasm. A number of *in vitro* techniques has been developed to refold the protein to the native (active) conformation (review: Middelberg 2002). Invariably this involves initially dissolving the inclusion bodies in a denaturing environment (urea or guanidine) and then removing the denaturant under conditions that promote correct refolding. The hydrophobic external faces of membrane proteins make these particularly vulnerable to aggregation and generally require the presence of detergent or lipid-type molecules in the renaturing medium. Using this technique successful results have been obtained, even for proteins with extensive transmembrane helices (Kiefer, 2002) or β-barrels (Buchanan,1999).

A more recent development of the method is first to purify the denatured inclusion bodies chromatographically, then renature the proteins while still bound to the column (Vincent *et al.*, 2004; Lemercier *et al.*, 2003). It is believed that the chromatographic binding of the denatured protein molecules prevents re-aggregation as the denaturant is removed. This technique, of course, requires the fusion of a His-tag in the expression construct.

There is an advantage to refolding inclusion bodies over expression of soluble fusion products. This is that, particularly where high yield has been achieved, purification (adequate for crystal formation) can be performed using centrifugation alone, and thus it is possible to expresses the *native* protein (ie. without his-tag or other fusion construct elements).

Two approaches were taken to refolding AT611 and p32-718 inclusion bodies: dialysis, and rapid dilution. In each case various refolding conditions were imposed (temperature, pH, salt concentration etc.) and the results assayed for any trace NADH dehydrogenase activity.

4.4.1 Protein refolding by dialysis

Samples of the denatured inclusion bodies (in 6M guanidine or 8M urea) were prepared by the addition of KCl to a final concentration of 0.1M, and KPi to 20mM at either pH6.0 or pH8.0: a total of eight preparations, four each for AT611 and p32-718. Two 1ml aliquots were taken from each of these, and FAD added to a final concentration of either 10μM or 1mM.

Table 4.5 Dialysis refolding buffers

Buffers used in experiments to refold inclusion bodies dissolved in either 6M guanidine or 8M urea. In each case KPi, KCl and FAD were added to the samples to the same concentration as that of the target buffer.

Detergents in buffers 5-8: OG: octylglucoside; SDS: sodium dodecyl sulphate

Buffer	KPi	pH	KCl	FAD	OG	SDS
1	20mM	6.0	100mM	10μM	-	-
2	20mM	6.0	100mM	1μM	-	-
3	20mM	8.0	100mM	10μM	-	-
4	20mM	8.0	100mM	1μM	-	-
5	20mM	6.0	100mM	100μM	0.5%	-
6	20mM	6.0	100mM	100μM	0.5%	0.05%
7	20mM	8.0	100mM	100μM	0.5%	-
8	20mM	8.0	100mM	100μM	0.5%	0.05%

Initially four dialysis buffers were prepared (buffers 1 to 4 in table 4.5). Dialysis tubing (Sigma #D9277[1]), equilibriated overnight in 20mM KPi, was filled by pipette with 500μl of each sample and sealed in 50ml Falcon® tube containing the appropriate buffer (matching FAD concentration and pH). A total of 32 dialysis experiments were thus performed, sixteen at room temperature and sixteen at 4°C. These were assessed for successful refolding by taking a 50μl sample from the tubing and assaying for NADH dehydrogenase activity. The room temperature experiments were assessed after one hour and again after overnight incubation. All 4°C experiments were left overnight.

No activity could be detected in any of the samples. A precipitate was visible in each suggesting that the protein had re-aggregated as a result of dialysis.

[1] retains > 90% cytochrome c (12.4kDa) in solution over 10 hrs

As membrane proteins have been demonstrated to fold correctly in the presence of detergent or lipid, similar experiments were performed with additional octylglucoside (OG) and SDS (buffers 5 to 8 in table 4.5).

Both room temperature and overnight 4°C experiments were performed for both constructs, but precipitation was again evident and no activity could be detected.

4.4.2 Protein refolding by rapid dilution

Although the controlled removal of denaturant (urea or guanidine) by the dialysis technique (above) may promote native refolding, the possibility exists that transient conditions may result in re-aggregation of the protein molecules. An alternative method

Table 4.6 Dilution refolding buffers

Buffer compositions for injection refolding experiments derived from Hampton Research FoldIt specification.

Notes: EDTA 1mM; Mg: 22mM $MgCl_2$, Ca: 2.2mM $CaCl_2$
PEG-2000 0.055% (w/v), Gaunidine-HCl 550mM,
Sucrose 440mM, L-Arginine 550mM

#	buffer	pH	FAD	NaCl	KCl	Chelator	Additives
1	55mM Tris	8.2	20μM	246mM	11mM	EDTA	PEG
2	55mM MES	6.5	20μM	10.56mM	0.44mM	Mg, Ca	Guanidine
3	55mM MES	6.5	20μM	10.56mM	0.44mM	EDTA	PEG, Guanidine, Sucrose, L-Arginine
4	55mM MES	6.5	20μM	246mM	11mM	Mg, Ca	Sucrose, L-Arginine
5	55mM MES	6.5	20μM	246mM	11mM	Mg, Ca	Sucrose
6	55mM Tris	8.2	20μM	10.56mM	0.44mM	EDTA	PEG, Guanidine, Sucrose
7	55mM Tris	8.2	20μM	10.56mM	0.44mM	Mg, Ca	Guanidine, L- Arginine
8	55mM MES	6.5	20μM	246mM	11mM	EDTA	PEG, L-Arginine
9	55mM MES	6.5	20μM	246mM	11mM	Mg, Ca	PEG, Guanidine, Sucrose
10	55mM Tris	8.2	20μM	10.56mM	0.44mM	EDTA	Sucrose
11	55mM Tris	8.2	20μM	10.56mM	0.44mM	Mg, Ca	Guanidine
12	55mM MES	6.5	20μM	246mM	11mM	EDTA	Guanidine, L-Arginine
13	55mM Tris	8.2	20μM	246mM	11mM	EDTA	Guanidine
14	55mM MES	6.5	20μM	10.56mM	0.44mM	Mg, Ca	PEG
15	55mM MES	6.5	20μM	10.56mM	0.44mM	EDTA	Sucrose, L-Arginine
16	55mM Tris	8.2	20μM	246mM	11mM	Mg, Ca	PEG, Guanidine, Sucrose, L-Arginine

subjects the denatured protein to rapid dilution (diffusion) into the target buffer. The choice of buffer/pH and concentration of additives (salt, redox poise agents, bulking molecules etc) is pivotal to attaining successful refolding, and a range of buffer compositions was employed. Commercial kits are available (Novagen Protein Refolding Kit #70123-3, AthenaES Refolding Kit, Hampton Research FoldIt Screen, and others). Sixteen buffers solutions were prepared using the FoldIt recommended compositions as detailed (table 4.6). 50µl denatured protein (10mg/ml) was injected swiftly by pipette into 1ml buffer (a 20-fold dilution, final concentration: 0.5 mg/ml), vortexed briefly and left to incubate.

192 dilution experiments were performed: four denatured protein samples (AT611 and p32-718, each in 6M Guanidine and 8M Urea), each with all of the following incubation conditions: 4°C / 4 hours, 4°C / 16 hours, and RT / 4 hours.

Table 4.7 Dilution Refolding buffer supplements

Additional buffer constituents used in refolding experiments. The numeric identifier for each refers to the buffer defined in table 4.6 to which the indicated compounds have been added at the concentrations shown.

buffer	NADH	DTT	Lauryl Maltoside	GSH	GSSG
1A	100µM	1mM			
2A	100µM		3mM	1mM	100nM
3A	100µM			1mM	100nM
4A	100µM	1mM	3mM		
5A	100µM			1mM	100nM
6A	100µM	1mM	3mM		
7A	100µM	1mM			
8A	100µM		3mM	1mM	100nM
9A	100µM	1mM			
10A	100µM		3mM	1mM	100nM
11A	100µM			1mM	100nM
12A	100µM	1mM	3mM		
13A	100µM			1mM	100nM
14A	100µM	1mM	3mM		
15A	100µM	1mM			
16A	100µM		3mM	1mM	100nM

A further 16 buffers were then prepared (buffers1A to 16A, table 4.7) by additions to the existing solutions. These included 'ligand' (NADH), redox couple, and detergent.128 further injection experiments were thus performed, using 4°C / 4 hours and RT / 4 hours as incubation conditions.

All 320 experiments were assayed for enzyme activity as described above, using 100µl in 1ml cuvettes (final protein concentration 50µg/ml). Additional controls included 100µl of each buffer. An enzyme rate of 2X the buffer control rate was deemed evidence of enzyme activity and would be use as a starting point for buffer optimisation. No enzyme activity was detected in any of these experiments.

4.5 Isolation of *Arum maculatum* NDH2

Arum maculatum (cuckoo-pint) is known to express elevated NDH2 activity, associated with the "rotting material" smell used to attract insects (Wagner *et al.*, 1988 & refs. therein). Specifically, the energy released by NDH2 activity causes a rise of over 10°C (above ambient) within the appendix. Isolation of NDH2 enzyme from this species serves two purposes in the context of this research: firstly it allows a degree of enzyme characterisation using biochemical analysis; and secondly N-terminal sequencing (by Edman degradation) may determine the mature protein terminus. The latter is particularly beneficial as the mature *Arabidopsis* protein termini may be inferred from this, and thus more suitable primers developed for recombinant expression.

Preparations of disrupted mitochondrial particles ("AM4") have previously been isolated in this lab (Prof P. Rich). These were obtained from spadices of *Arum maculatum* plants immediately prior to the onset of thermogenesis phase and stored at –80°C. Isolation of NADH dehydrogenase activity from the external face of *Arum* mitochondria has been achieved by deoxycholate solubilisation followed by chromatographic purification (Cottingham & Moore, 1984). In this report bands of 65kDa and 78kDa were observed on SDS-PAGE gels.

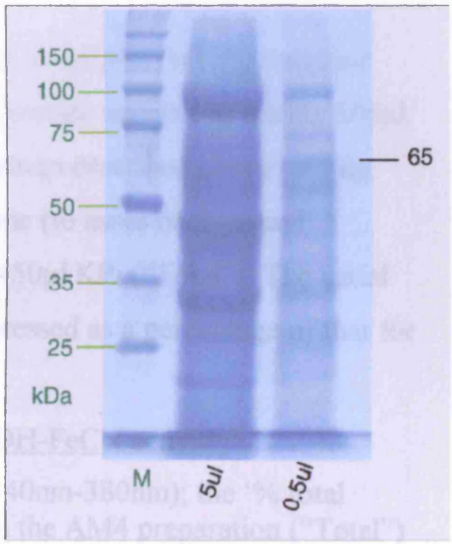


Fig 4.15 SDS-PAGE analysis of *Arum maculatum* protein composition
in AM4 mitochondrial preparation.

5µl = 300µg protein, 0.5µl = 30µg protein, see text.

M: mol. wt. markers (5µl, Promega # V8491

Initial analysis using the Bradford assay revealed a protein concentration of 60mg/ml in the AM4 preparation, and an SDS-PAGE gel (using 5µl and 0.5µl samples) exhibited a putative NDH2 band at around 65kDa (fig 4.15) consistent with that previously reported.

A more crude isolation of the enzyme was attempted by detergent solubilisation followed by salt precipitation, using the NADH-FeCN time-course assay to assess enzyme activity in the various fractions. Five detergents were selected as candidates for

NDH2 solubilisation: EDTA, Cholate, Tween-80, Octyl-glucoside and MEGA-8. Solutions of these were prepared in 0.1M potassium phosphate buffer (KPi) pH8.0. A 50µl sample of AM4 was first suspended in 500µl KPi buffer alone, vortexed and then centrifuged for 30 minutes at 300,000g (S100-AT4 ultracentrifuge rotor, 99,000 rpm, 4°C). The (mitochondrial membrane) pellet was then resuspended in 500µl KPi. Analysis of the NADH dehydrogenase activity revealed roughly equal values for the membrane and AN4 soluble fractions. Whether the unexpectedly high soluble activity was attributable to dissociated NDH2 or other soluble dehydrogenase enzymes could not be determined; however an assay using KPi buffer alone produced minimal background activity thus eliminated it as a contributor. Consequently the AN4 preparation was centrifuged and resuspended three times in an equal volume of KPi buffer (using the above protocol) to isolate membrane associated activity.

50µl of the washed AM4 was added to 450µl of each the detergent buffers, vortexed briefly, centrifuged and the supernatant dehydrogenase activity assayed by adding 100µl of detergent solution to the NADH/FeCN/buffer reagents as described above (see fig 4.16). Similar assays were performed for KPi buffer alone (to asses background activity), and for a suspension of 50µl washed AM4 in 450µl KPi ("Total"). The initial absorption rate for each was tabulated (fig 4.8) and expressed as a percentage of that for

Table 4.8 Detergents' effectiveness at solubilising NADH-FeCN activity.

dA/dt represents the measured change in absorbance (340nm-380nm); the '% total activity' is calculated with respect to that obtained from the AM4 preparation ("Total") after correction for background activity in the presence of buffer alone. **Bold figures** indicate highest activity values.

Detergent	concentration	dA/dt ($As^{-1}*10^{-3}$)	corrected	% of total
None (KPi buffer alone)		0.2		
Total(AM4 unsolubilised)		12.3	12.1	100
EDTA	0.1M	2.8	2.6	22
Cholate	0.1% (w/v)	4.3	4.1	**34**
Tween-80	1% (w/v)	5.9	5.7	**47**
Octyl-glucoside	10mM	2.4	2.2	18
MEGA-8	10mM	2.6	2.4	20

the resuspended pellet. Cholate and Tween-80 proved superior to the other detergents in this assay.

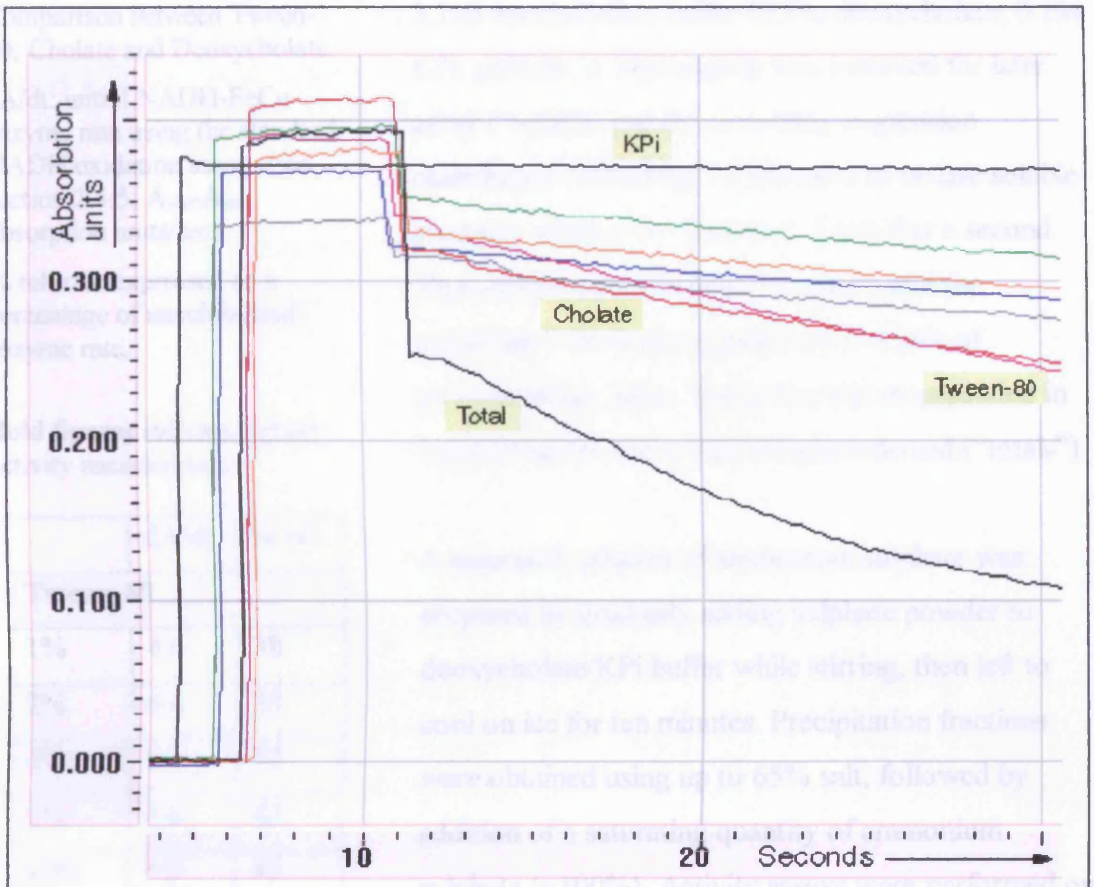


Fig 4.16. Dehydrogenase activity of detergent solubilised AM4

Enzyme activity detected by rate of NADH oxidation (absorbance at 340nm-380nm)

KPi: control - no enzyme or detergent
Total: 50µl washed AM4 in 450µl KPi buffer.
Cholate (0.1% w/v) and Tween-80 (1% w/v) traces marked

A further assay was performed to determine the optimum detergent concentration using the washed AM4; this included the related detergent, deoxycholate, at the concentrations shown (table 4.9). From the results of this assay it was decided to use 0.5% deoxycholate to solubilise the mitochondrial particle suspension.

Table 4.9 Assay for optimum detergent concentration

Comparison between Tween-80, Cholate and Deoxycholate.

dA/dt: initial NADH-FeCn enzyme rate using the standard NADH oxidation assay (see section 2.4.5; A_{340}-A_{380} absorption units/sec)

% rel: rate expressed as a percentage of unsolubilised enzyme rate.

Bold figures indicate highest activity measurement

	dA/dt	% rel
Tween-80		
1%	4.6	38
2%	6.4	53
4%	3.7	31
6%	4.8	40
10%	5.2	43
Cholate		
0.25%	5.2	43
1%	7.0	58
2%	5.3	44
3%	3.7	30
6%	3.4	28
Deoxycholate		
0.25%	7.5	**62**
1%	6.0	50
2%	7.1	51
3%	4.0	33
6%	3.7	31

310µl stock AM4 was centrifuged at 100,000g for 10 minutes and the pellet suspended by vortexing in 3.1ml deoxycholate buffer (0.5% deoxycholate, 0.1M KPi, pH8.0). A 50µl sample was removed for later assay ("total"), and the remaining suspension centrifuged (300,000g, 30 minutes) to isolate soluble products which were decanted. From this a second 50µl sample was removed ("$total_{SN}$") and the remaining 3ml sample transferred to a pair of ultracentrifuge tubes. The pellet was resuspended in 3ml KPi buffer and a 50µl sample removed ("$total_P$")

A saturated solution of ammonium sulphate was prepared by gradually adding sulphate powder to deoxycholate/KPi buffer while stirring, then left to cool on ice for ten minutes. Precipitation fractions were obtained using up to 65% salt, followed by addition of a saturating quantity of ammonium sulphate (=100%). Activity assays were performed on the pellet and supernatant samples from each stage (table 4.10). These data do not suggest that the activity was precipitated in a single (or pair) of precipitation fractions. It is also curious that significant activity was still present in the supernatant after addition of a saturating salt concentration.

As it is possible that sulphate may impede NADH dehydrogenase activity a second precipitation experiment was performed using ammonium acetate. However this produced similar results so also failed to isolate dehydrogenase activity to any precipitation fraction.

Table 4.10 Ammonium sulphate precipitation of NADH dehydrogenase activity

dA/dt: initial NADH-FeCn enzyme rate using the standard NADH oxidation assay (see section 2.4.5: A_{340}-A_{380} absorption units/sec). Values shown are those for both the supernatant and resuspended pellet for each salt addition step.

Salt (%)	Solution (dA/dt)	Precipitate (dA/dt)
0	3.62	
20	2.88	0.93
25	2.56	0.33
30	2.31	0.31
35	2.30	0.33
40	2.27	0.38
45	2.28	0.35
50	1.75	0.38
55	1.44	0.47
60	1.27	0.30
65	1.02	0.31
100	0.74	0.64

As NDH2 dehydrogenase activity in purified protein is known to be unstable both in *Arum maculatum* (Cottingham & Moore, 1984) and other species (Björklöf *et al.*, 2000), whether as a consequence of detergent concentration, requirement for the presence of lipid, protein degradation or a combination of these, it was decided that the results potentially obtainable did not justify further pursuit of this avenue of investigation.

4.5 Summary of expression and biochemical investigation

Restriction digest analysis of the inherited *Arabidopsis thaliana* NDH2 cDNA clones (AT1, AT2, AT4 and AT6) indicated that neither AT2 nor AT4 contained valid gene inserts. However, AT611/BL21 (AT1 subcloned into an expression vector and transformed into BL21 *E. coli* cells) was expressed, and SDS-PAGE analysis revealed an expressed protein band of the predicted mass. As an initial goal had been the comparative biochemical analysis of different *Arabidopsis* NDH2 enzymes it was decided to adopt parallel tasks of *de novo* construction of cDNA clones from *Arabidopsis* leaf material, and expression of **nda1** (ie AT6).

Arabidopsis thaliana plants (strain col-0) were grown from seed, their leaves harvested and RNA was purified from these using the Qiagen Plant RNeasy system. A successful purification was concluded through quantification of RNA concentration and agarose gel analysis for RNA integrity. Forward and reverse primers were designed for the NDH2 genes incorporating suitable restriction sites for purpose of subsequent cloning. RT-PCR using Qiagen Sensiscript and Vent polymerase failed to produce any products, but a putative **nda1** product (as determined by agarose gel electrophoresis) was obtained when substituting Quiagen Sensiscript (RT) and Taq polymerase (PCR) enzymes.

As **nda1** is the most abundant NDH2 mRNA species (Michalecka, 2005) it is not surprising that a sub-optimum RT-PCR protocol will yield this transcript alone. With the aim of acquiring the full complement of NDH2 cDNA clones alternative protocols were examined. The optimum solution was first to purify mRNA from the RNA sample using the Quiagen Oligotex system (mRNA is typically just 3% of total RNA), then to use a 1:1 mixture of Sensiscript and Omniscript in the RT reactions with 200ng mRNA template (four times that recommended by the manufacturer). This produced convincing additional putative PCR products (as determined by agarose gel electrophoresis) for **ndc1** ('874') and **ndb1** ('822'), a likely product for **nda2** ('999'), as well as a possible **ndb2** product ('502'). Further refinements to the protocol (eg. annealing temperature) is likely result in the systhesis of the complete NDH2 cDNA set, though this was not

deemed prudent to perfect this in the current work prior to successful expression of **nda1**.

Early attempts to express AT611 (**nda1**) showed that the protein was (at least, predominantly) aggregating as inclusion bodies - both by SDS-PAGE analysis of cell fractions and the visible aggregation of precipitate directly following cell disruption. As NDH2 is believed to be a membrane surface-associated enzyme (rather than membrane-embedded) it was hoped that expression of the gene fused to a soluble domain would render a soluble (or partly-soluble) protein product. Consequently the gene was subcloned (with alterations) into the pET32 vector to provide the following potential benfits: 1) protein fused to a soluble Trx domain, 2) N-terminal fusion, as NDH2 C-termini are more conserved and thus may be internal to the protein, 3) removal of the mitochondrial targetting sequence which serves no purpose in a bacterial expression host and may impede solubility, and 4) incorporation of an enterokinase site allowing cleavage of the NDH2 product following expression.

Attempts to express this construct (p32-718) and AT611 under differring expression conditions, however, failed to result in any detectable NADH-FeCn oxidoreductase activity in the soluble fractions of these experiments. It was thus decided to prepare samples of each product by dissolving the inclusion body fractions in guanidine or urea and attempt to refold these by the techniques of dialysis and rapid dilution. These experiments were performed under varying conditions of temperature, pH and additives (eg. detergent) and in each case FAD (the NDH2 cofactor) was included in the reaction. However in none of the 64 dialysis and over 300 injection experiments could any NADH-FeCn oxidoreductase activity be detected in the resulting solutions.

In order (principally) to determine the mature N-terminus of a plant NDH2 by Edman degradation, AN4 (a preparation of disrupted mitchondria from the spadices of *Arum maculatum*) was subjected to detergent solubilisation followed by salt precipitation in the hope of obtaining purified NDH2 (as detected by NADH-FeCn oxidoreductase activity. This has previously been achieved using deoxycholate solubilisation followed by chromatographic purification (Cottingham & Moore, 1984). From a number of detergents examined deoxycholate proved the optimum detergent as determined by the NADH-FeCn activity assay. The mitochondrial preparation (AN4) was thus dissolved in

0.5% deoxycholate and precipitation experiments performed using using increasing concentrations of ammonium sulphate and ammonium acetate. However, these experiments failed to isolate NADH-FeCn activity to any one (or pair of) precipitation fraction(s).

Chapter 5:
In silico investigation into NDH2 structural features

5.1 Introduction

Currently no NDH2 structure has been resolved by NMR or X-ray crystallography. However a model for the *E. coli* enzyme has been published (Schmid & Gerloff, 2004) using the known structure of *Streptococcus faecalis* NADH peroxidase as the template (PBD: 2NPX). As the sequence homology between *E. coli* and the *Arabidopsis* NDH2s is significant (Chapter 3), models of the latter using the former as a template are likely to represent useful approximations of the physiological structures. From these it is possible to elucidate key features of these enzymes, assess functional implications of conserved primary structure motifs, and make predictions such as NADH/NADPH specificity.

5.2 Secondary structure prediction of NDH2 genes

A list of eighteen NDH2 genes, selected to represent the species-wide diverse groups identified in Chaper 3, was assembled to create a 'representative' NDH2 group. (This includes two *Trypanosoma* members which, while failing to qualify by the stated criteria, exhibit close homology to **nda** genes. These have C-terminal additional domains of 55-75 residues.)

The first step in this analysis was the prediction of secondary structure for these eighteen representative NDH2s.
To assess the relative suitability of web-based prediction servers for NDH2 enzymes the amino acid sequence of 2NPX was submitted to four net servers (NNPREDICT, SOMPA, NPSA and jpred[1]) and the results compared against the X-ray crystallographically determined secondary structure. The latter was extracted from the

[1] SOMPA: www.sompa; NNPREDICT: http://www.cmpharm.ucsf.edu/~nomi/nnpredict.html NPSA: http://npsa-pbil.ibcp.fr/NPSA jpred: http://www.compbio.dundee.ac.uk/~www-jpred/

PDB file by loading into the Swiss PdbViewer[1] and manually annotating the sequence. Jpred (which compiles a consensus derived from various prediction algorithms) produced the best matching results for 2NPX and was thus selected for all NDH2 prediction.

The eighteen sequences were annotated for presence of predicted alpha helices and beta strands and aligned accordingly (fig. 5.1). Reassuringly there is a high degree of secondary structure homology across all selected NDH2s, even where there is significant primary structure divergence. This implies that it should be possible to model each from the *E. coli* template with some confidence. It should be noted though that there is no significant secondary structure homology in the N-terminal regions of the eukaryotic proteins, lending weight to the argument that these are a target signal sequences possibly cleaved to the prokaryotic N-terminus locus in the mature proteins (Chapter 3).

[1] version 3.7 http://www.expasy.org/spdbv (aka DeepView, abbreviated 'Spdbv' in this chapter)

Fig 5.1 Alignment of Representative NDH2s. **(1/3)**

Secondary structure predicted by **jpred** (RED: Alpha helix, BLUE: beta strand)

AtI: *Arabidopsis thaliana* nda1 (NM_100592); OsI: *Oryza sativa* (AP003705); Tbruc: *Trypanosoma brucei* (EAN78339); Tcruz: *Trypanosoma cruz* (EAN90346); Nc FI: *Neurospora crassa* (XM_331371); Sc FI *Saccharomyces cerevisiae* (SCYDL085W); AtII *Arabidopsis thaliana* ndb1 (AL161572); OSII: *Oryza sativa* (AK102878); NcFII: *Neurospora crassa* (NCR236906); AtIII: *Arabidopsis thaliana* ndc (NM_120955); nos: *Nostoc sp.* PCC 7120 (AP003595); ba: *Bacillus anthracis* (NC_003995); ec: *Escherichia coli* (NC_00091).

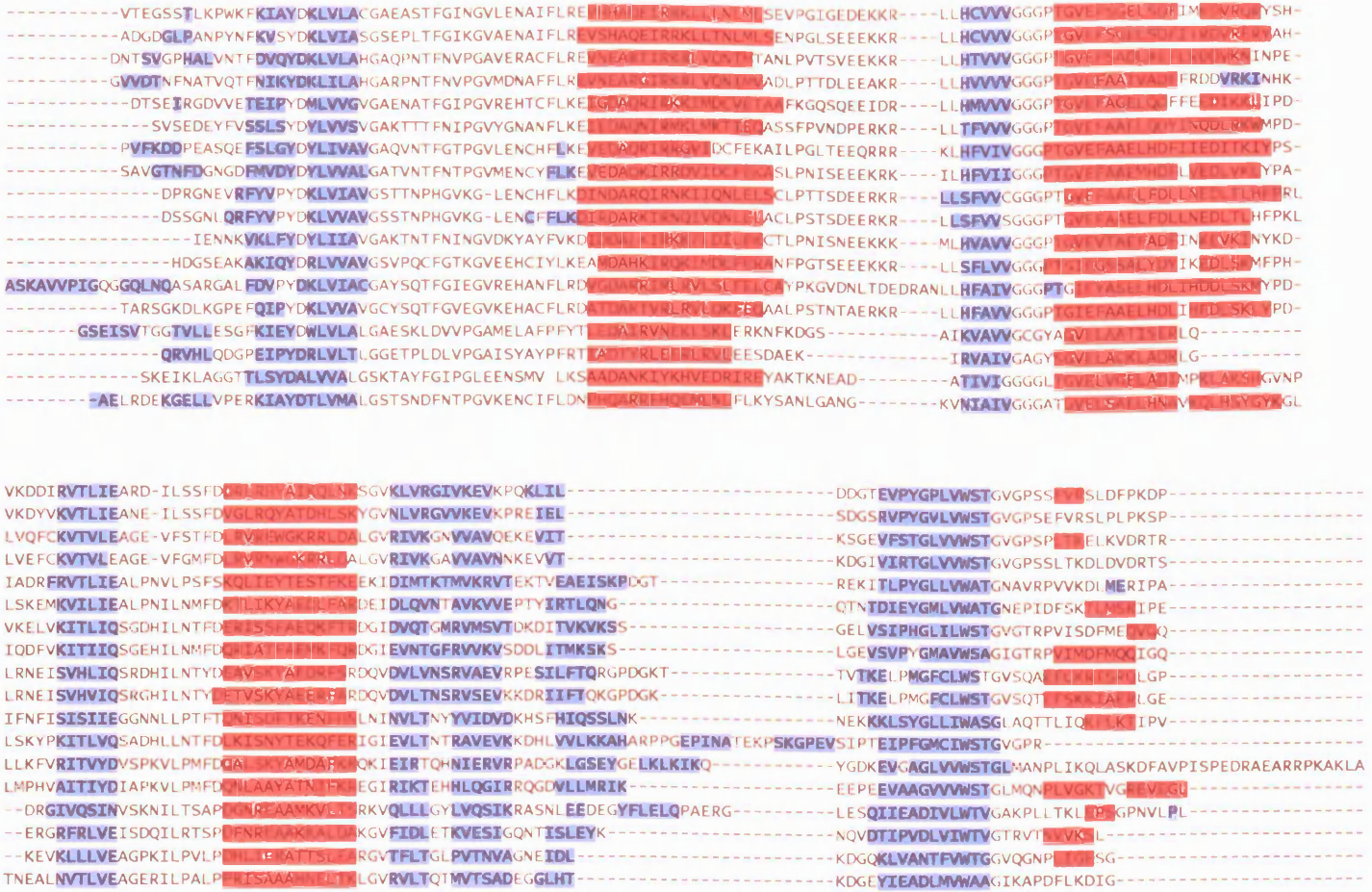


Fig 5.1 (2/3)

Fig 5.1 (3/3)

5.3 Review of 1OZK

The PDB file for 1OZK, the model created for *E. coli* NDH2 (Schmid & Gerloff, 2004), was obtained[1] and loaded into Spdbv. The secondary structure was extracted (as described above for 2NPX) and compared to that predicted by jpred: generally the agreement was close but one predicted 12-residue helix (α_5 in fig 5.3) was absent in 1OZK. This was unexpected as the predicted helix is common to all NDH2s in spite of sequence diversity at this locus. Examination of the 3D structure in Spdbv revealed an unstructured loop on the external face of the protein.

The model was examined in more detail particularly, in the region of the FAD isoalloxazine moiety which lies at the heart of the enzyme. A further unexpected discovery was the presence of a leucine residue (L49) obscuring the postulated quinone site (described below).

It was decided to remodel a structure for *E. coli* NDH2 in order to compare this with the 1OZK solution, and possibly draw alternative conclusions as to structural motifs and functional aspects of the enzyme.

[1] http://oca.ebi.ac.uk/oca-bin/ocashort?id=1ozk (see also Supplementary Data)

5.4 Computational synthesis of ecNDH2

5.4.1 Identification of potential templates

The protein sequences for each of the 18 representative NDH2s were submitted to the NCBI and T-02[1] servers for identification of homologous structures in the PDB database using two iterations of PSI-BLAST and the results pooled. Most of the high-scoring homologues, including 2NPX, were classified in SCOP[2] as members of the "FAD/NAD-linked reductase" family (review: Dym & Eisenberg, 2001). Consequently all SCOP-classified proteins were selected and the deposited structures downloaded from the Rutgers PDB (Brookhaven) database[3] (both PDB files and Fasta-format sequences) with additional supporting documentation from the EBI website[4]. One structure from each subfamily was then selected for closer examination based on the inclusion of substrates in crystal structure (where available) and excluding engineered proteins, with three additional structures not listed in the SCOP family but identified by PSI-BLAST (table 5.1).

As expected, homology for the NDH2 C-terminal domains with the selected structures was poor, so the 60 terminal residues of each was used for a further PSI-BLAST search of the PDB database but this too failed to identify putative homologues.
Finally the EBI site was searched (text search) for structures of enzymes with bound quinone (or synthetic homologues) and a number of files downloaded for later examination.

[1] http://www.soe.ucsc.edu/research/compbio/HMM-apps/T02-query.html
[2] http://scop.mrc-lmb.cam.ac.uk/
[3] http://pdbbeta.rcsb.org/
[4] http://ca.ebi.ac.uk

	Subfamily name	PDB	notes
1	NADH-dependent 2-ketopropyl CoM oxidoreductase/carboxylase	**1MOK**	3
2	Dihydrolipoamide dehydrogenase	**1LVL**	N
3	Glutathione reductase	**1GET**	N
4	Mercuric Reductase	**1ZK7**	1
5	Trypanothione reductase	**1AOG**	M
6	OM protein p64	**1BHY**	2
7	NADH-dependent ferredoxin reductase	**1F3P**	N
8	Putidaredoxin reductase	**1Q1R**	
9	Apoptosis-inducing factor (AIF)	**1GV4**	
10	Flavocytochrome c sulfide dehydrogenase [subunit]	**1FCD**	1, H
11	Nitrite Reductase	**1XHC**	1
12	NADH peroxidase	**2NPX**	N
13	Thioredoxin reductase	**1VDC**	p
14	Alkyl hydroperoxide reductase [subunit - C-term domain]	**1FL2**	(p)
15	Phenylacetone monooxygenase	**1W4X**	

Table 5.1 Representative FAD/NADH-linked oxidoreductases

N: bound NAD(P)H; M: bound MAE; H: Heme bound
p: NADP substrate (p): NADPH or NADH substrate
1: not listed in SCOP family; 2 (membrane dihydrolipoamide dehydrogenase)
3: see also **1MO9** which has bound CoM

5.4.2 Structural analysis of FAD/NADH-linked oxido-reductases

The fifteen selected homologous oxido-reductase structures were loaded into Spdbv as before in order to compare their tertiary structure and extract secondary structure information. The latter was used manually to annotate sequence data which was aligned (as far as possible) by tertiary homology (fig 5.3). Included in this figure are the jpred-predicted secondary structure for *E. coli* NDH2 ("ec[jpred]") and the secondary structure extracted from the 1OZK model ("ec[1OZK]").

Comparison of these data proved instructive, allowing unambiguous identification of conserved structural features around the FAD and NAD(P)H sites and diversity in the tertiary structure around the second substrate locus (and thus primary structure loci).

1MOK is distinct from the other enzymes in that it has an additional 40-residue N-terminal domain, while 1VDC and 1FL2 lack the C-terminal dimerisation domain (in spite of being annotated as homodimers) Uniquely among this group 1W4X terminates near β_{11}.

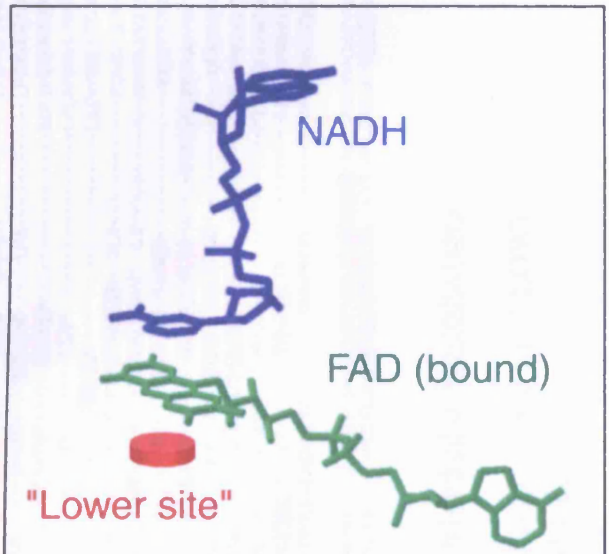


Fig 5.2 Core components of the electron transfer pathway

At the heart of each enzyme is the bound FAD molecule; where the crystal includes the NADH substrate this is orientated with its nicotinamide ring parallel to, and around 3Å 'above', the FAD isoalloxazine ring (but see exception below). Beneath this is a region which is accessible (to a greater or lesser extent) from the lower side of the enzyme, which may termed the "lower site" (fig 5.2).

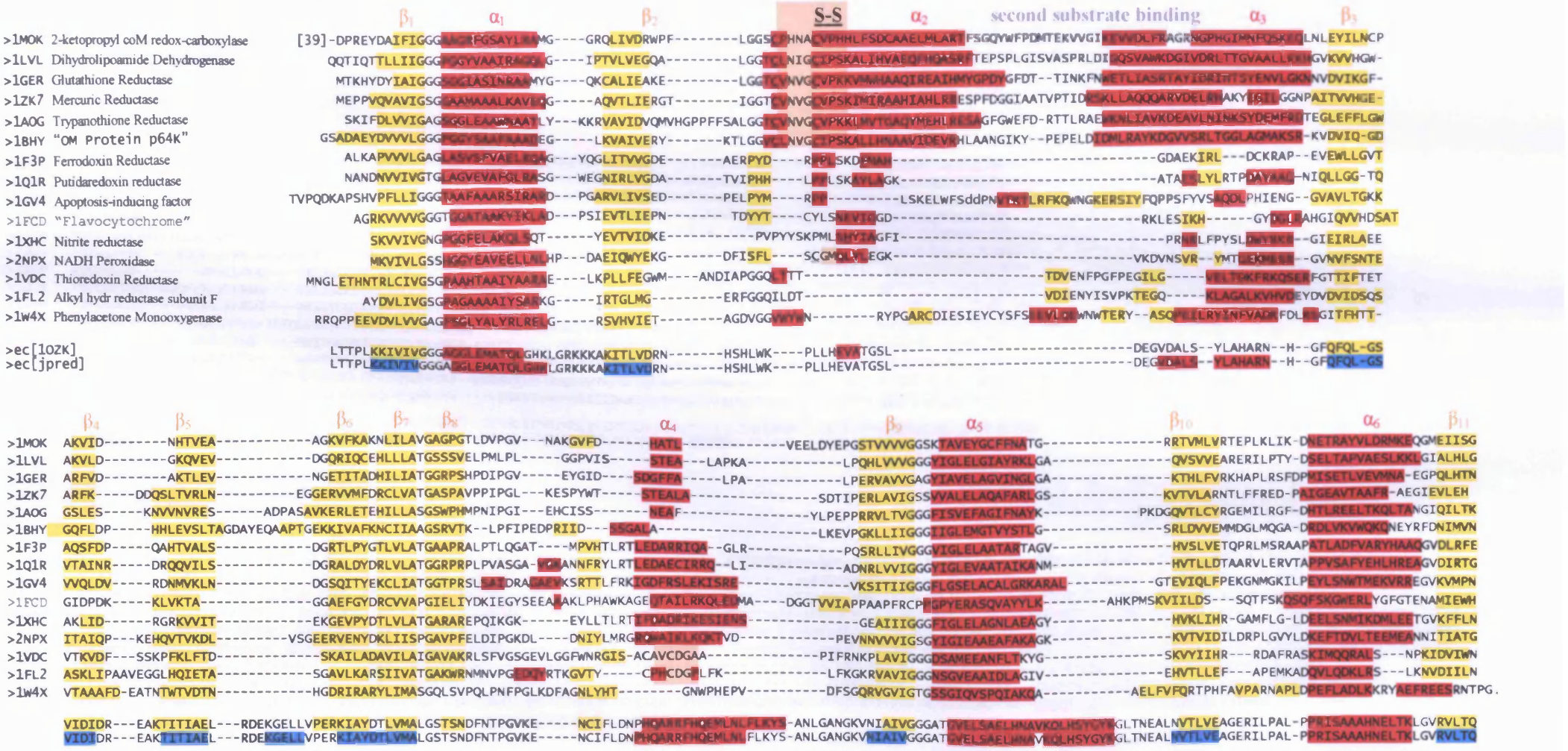


Fig 5.3 (1/2) Alignment of Representative FAD/NADH-linked oxido-reductases

Annotated 3D-determined secondary structure:helix (RED), strand (YELLOW)
Predicted secondary structure:helix (RED), strand (BLUE)
lower case: residues not included in structure
Pink boxes: cysteine redox sites

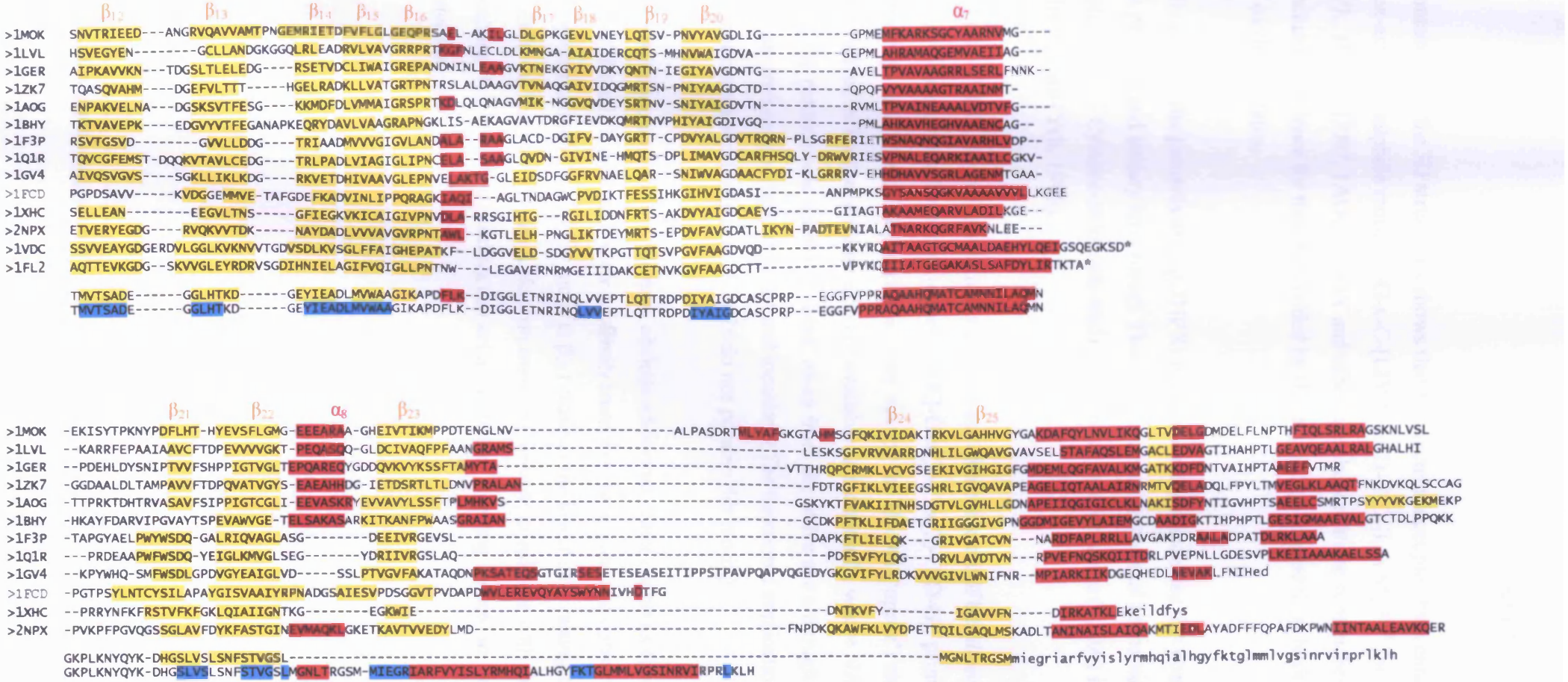


Fig 5.3 (2/2)

Examination of the 3D structures shows that this locus is occupied by a canonical redox-active disulphide motif (G-G-x-C-[LIVA]-x(2)-G-C-[LIVM]-P)[1] for 1MOK, 1LVL, 1GET, 1ZK7, 1AOG, 1BHY and other members of these subfamilies. Conclusive evidence for this is provided by the deposition of models in both reduced and oxidized states.

In the case of the peroxidases (eg 2NPX) a single cysteine occupies this position (in both primary and tertiary structures). This too serves as an unusual redox site (Cysteine ⟷ Cysteine-sulphenic acid) with one crystal structure (1JOA) in the oxidised state (Yeh, 1996).

1FL2 and 1VDC and their subfamilies also have a disulphide motif but these differ from the previous examples both in sequence ([RK]-G-x(3)-C-x(2)-C-D-G), primary locus (between β_8 and β_9) and 3D position as they occupy the NADH ('upper') site. Again, both oxidized and reduced structures are available and, curiously, where NADH is present the nicotinamide moiety is 'bent' away from the active site although the remaining molecule occupies the normal location. (The functional implications of this were not considered further as NDH2s do not possess this motif.)

The remaining enzyme types have no obvious additional redox elements except that 1F3P, 1QIR and 1GV4 (and their subfamily members) have identical conserved tryptophan residues (in a WSD motif at β_{21}) close to the lower site, raising the possibility of involvement in an electron transfer pathway, consistent with the hypothesis that aromatic residues can act as transient 'stepping stones' in electrically conductive pathways.

[1] Prosite: www.expasy.org/prosite/

5.4.3 Modelling the *E. coli* NDH2

The initial chimeric model

Models were created using version 8.2 of MODELLER[1], a script-directed program (no graphical interface) which uses the python language with extensions. Inputs to the program are template structures (PDB files), sequence data (for the protein to be modelled), an alignment file (to direct the homology modelling) and a python script file linking these with, optionally, additional restrictions.

For the initial model the C-terminus was removed from the *E. coli* sequence (β_{23} and beyond) as this finds no obvious equivalence in the available templates and may compromise homology modelling. The C-terminal abbreviated protein was termed ecCx[2]:

```
>P1;ecCx
sequence:ecCx:::::::0.00:0.00
LTTPLKKIVIVGGGAGGLEMATQLGHKLGRKKKAKITLVDRNHSHLWKPLLHEVATGSLDEGVDALSYLAHARNHGFQ
FQLGSVIDIDREAKTITIAELRDEKGELLVPERKIAYDTLVMALGSTSNDFNTPGVKENCIFLDNPHQARRFHQEMLN
LFLKYSANLGANGKVNIAIVGGGATGVELSAELHNAVKQLHSYGYKGLTNEALNVTLVEAGERILPALPPRISAAAHN
ELTKLGVRVLTQTMVTSADEGGLHTKDGEYIEADLMVWAAGIKAPDFLKDIGGLETNRINQLVVEPTLQTTRDPDIYA
IGDCASCPRPEGGFVPPRAQAAHQMATCAMNNILAQMNGKPLKNYQYKDHGSLVSLSNFSTVGSLMGNLTRGSM*
```

There is no evidence of cysteine redox sites in the *E. coli* NDH2 sequence, nor of the β_{21} tryptophan motif, supporting the postulation that ubiquinone interacts directly with the FAD isoalloxazine ring (Schmid & Gerloff, 2004). The template exhibiting the best overall homology, both from PSI-BLAST results and by visual inspection, is 1XHC. However for the critical secondary substrate binding site region (β_2 to β_3) 1FL2 is a more suitable candidates, partly because its secondary structure is closer to that of ecCx, but more importantly because it has an exposed lower face of the FAD isoalloxazine ring - presumably to allow direct access by the second substrate as is assumed to be the case for NDH2s.

[1] http://salilab.org/modeller/
[2] this and all modelling source and structure files are contained on the Supplementary Data CD

MODELLER allows 'chimeric' constructions by providing alignment information from two non-overlapping (or partially overlapping) templates, provided the templates have the same orientation. This approach was attempted using a four residue overlap between 1FL2 and 1XHC at β_8 where the homology between templates is virtually perfect (see fig 5.3).

In order to obtain a correct template relative orientation, 1FL2 was manually positioned within Spdbv to superimpose it upon 1XHC (at the β_8 and FAD/NADH loci, where misalignment could be restricted to 0.2Å). There is a secondary structure element, a β-sheet, which now has strand components in both templates (β_7 in1FL2 and β_{20} in 1XHC); this too was within 0.2Å of superposition. The reoriented template was saved ('1FL2fitXH') and a model created from the now aligned pair (1XHC and 1FL2fitXH).

Visual inspection of the model showed that the synthesised chimeric domains were substantially misaligned/misorientated in the resulting models, a situation which could not be improved by modification to the alignment locus and/or overlap. The problem appears to be related to the format in which Spdbv saves PDB files: atom coordinates are recomputed but apparently to an internal coordinate frame to which an Spdbv-specific rotation matrix is appended. When reloaded to Spdbv the realigned model is correctly displayed, but MODELLER (ignoring the Spdbv matrix information in its computations) creates two linked but consequently misorientated domains.

The solution which resolved this problem was an unorthodox (and possibly novel) approach to protein modelling: creating a 'chimeric template' within Spdbv composed of suitably aligned peptide and ligand elements from different source templates, then creating a new composite molecule (**Edit→Create Merged Layer from Selection**). In spite of the fact that the resulting PDB file retains the original residue numbering and has disconnected peptide units MODELLER does not reject either of these artefacts[1].

[1] It is necessary also to employ multiple copies of each template which contributes more than one peptide element, load templates into Spdbv in the order of the final construct, and delete all templates before saving the merged layer in order to achieve a Spdbv generated PDB file acceptable to MODELLER.

The initial 1FL2-1XHC chimeric template was created with a junction at β_5. This produced an acceptable initial model from the hybrid template exhibiting no anomalies at the peptides' junction (see Supplementary Data):

5.4.4 NADH 'ligand'

Ligand interactions can provide valuable restraint parameters for MODELLER but neither 1FL2 nor 1XHC was crystallised with bound NADH (a 'ligand' in the modelling sense). However there are nine structures that have been crystallised with bound NAD(P)H (table 5.2), though as 1GEU has been engineered at NADH-binding residues it is discarded here.

PDB	substrate	FAD
2NPX	NADH	π
1F3P	NADH	π
1GEU[E]	NADH	π
1ZMD	NADH	π
1LVL	NADH	x
1GRB	NADPH	π
1GET	NADPH	π
1TYP	NADPH	x
1H6V	NADPH	-

Table 5.2 Structures incorporating NAD(P)H and FAD-relative orientation

nicotinamide moiety:
π : 'stacked' to FAD isoalloxazine ring
x : 'bent away' from FAD ring;
- : not defined in structure

[E] : Engineered gene; PINK: NADPH

These were examined in Spdbv to identify substrate-enzyme interactions. This revealed three loci on NADH (four on NADPH) (fig 5.4) but, unlike FAD, there is considerable variability in the conformation of the NAD(P)H molecule and the residues participating in H-bond interactions (table 5.3). In particular the nicotinamide ring in two structures is 'bent' 90 degrees away from the consensus position (parallel to the FAD isoalloxazine ring).

Whether this is an artefact of the crystallizing conditions or peculiar to these enzymes or their redox states is unknown. (Where enzyme structures are available in both reduced and oxidized states, such as 1VDC/1FL2 and 1NPX/1JOA, there is no evidence of a change in conformation at the NAD(P)H binding site.)

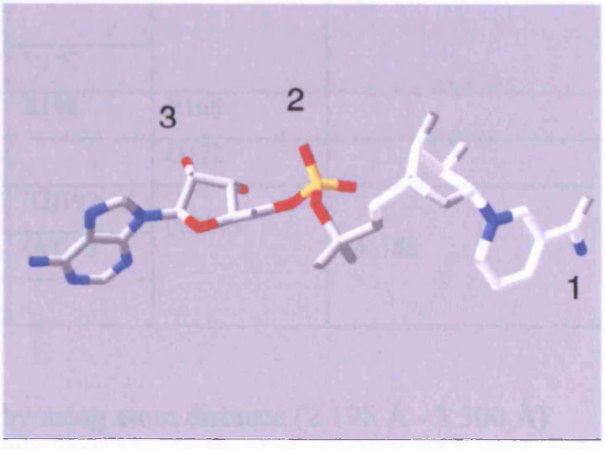


Fig 5.4 NADH H-bonding loci

NAD(P)H locus	PDB	residiue	Equivalent in family representative		Residue in 1XHC	Potential equivalent residue(s) in *E. coli*
1	1ZDM	E192	1LVL	E185	E149	E184
	1F3P	E159				
	1GEU	E181	1GET	E181		
	1GRB	E201	1GET	E181		
	1F3P	(W)320			(V)301	(G)363/(V)365
	1ZDM	(V)357	1LVL	(V)342		
	1GET	(V)342				
	1GEU	(V)342	1GET	(V)342		
	1GRB	(V)370	1GET	(V)342		
	2NPX	(G)328				
	{1LVL}	{R266}			{I224}	
2	1F3P	R183			G172	A223
	2NPX	Y188				
	1ZDM	(V)188	1LVL	(Y)181	(F)145	(G)182
P	1GET	R204			G172	P222
	1GET	R198			R166	A216
	1GRB	R218	1GET	R198		
	1TYP	R222	1AOG	R223		
	1TYP	Y221	1AOG	Y222	H165	E215
	1GRB	R224	1GET	R204	G172	P222
	1H6V	S222			G167	
	1H6V	R226			L171	
3	1LVL	E201[2]			H165	?
	1F3P	E175				
	1ZDM	E208[2]	1LVL	E201[2]		
	2NPX	D179				
	1GEU	E197	1GET	V197		
	1GRB	R218	1GET	R198	R166	
	1F3P	S182			G172	
	1TYP	(G)196	1AOG	(G)197	(G)143	(A)180
	1GRB	(A)195	1GET	(A)175		
	1H6V	(A)198				

Table 5.3 NAD(P)H binding sites

Hydrogen bond loci determined by Spdbv using atom distance (2.195 Å - 3.300 Å) NADPH bound structures are shaded pink. Residues in brackets () are backbone bonds. Braces [] indicate multiple bonds. Curved braces{} indicate binding to 'deformed' substrate. Residues incapable of equivalent H-bonding printed in grey.

It was deemed inappropriate to include NADH as a second 'ligand' principally because it might impose invalid constraints on NDH2 conformation, but also because none of the available NADH structures could be oriented to a near-perfect fit to 1XHC (less that 0.5 Å divergence at all potential binding sites).

4.4.5 Model Refinement

The iterative refinement technique employed here was comprised of:

- editing the template using sections from different FAD/NADH oxido-reductase structures,
- fine tuning the primary sequence alignment,
- placing restraints on MODELLER (eg. helix length), and
- introducing gaps in the alignment file to allow MODELLER increased flexibility.

Changes were made in response to:

- conserved (across NDH2s) secondary structure elements which differ from the template,
- unstructured loops (by visual inspection) in the model,
- primary structure elements present in *E. coli* NDH2 and other FAD/NADH oxido-reductases, though possibly absent in the template.

At each iteration the model was assessed visually for the effects of the changes.

For illustrative purposes the design process for one locus is described in detail, followed by a summary of the remaining loci.

5.4.5.1 Refining α_7

Jpred predicted a 21-residue α-helix starting PPRAQAA... which was aligned against the start of the homologous 1XHC helix (α_7). From this MODELLER constructed a 15-resiude α-helix (the length of the 1XHC helix) start-aligned as requested, followed by an unstructured loop.

There is virtually no sequence homology across the FAD/NADH family in this helix other than a conserved glutamine at the 7th residue[1] in four of the five enzymes which have a β-β mini-loop between β_{20} and α_7. Examination of the models shows that this residue forms a hydrogen bond with β_{20}. Although the jpred structure for *E. coli* NDH2 does not predict a β-β motif here, there *is* an extended length of residues and the helix has a glutamine residue in the (currently) 8th position. The resolution adopted was to model the loop region on 2NPX, offset the helix start in the alignment file by plus-one residue (to place the glutamine at position 7) and to direct MODELLER to construct a 20-residue helix from the new start position.

In order to model the loop FLXH.pdb was reloaded into Spdbv, then 2NPX.pdb and finally a second copy of FLXH.pdb. 2NPX was orientated to align with the FLXH models and two loci selected as edit points to minimise the gaps. A 20-residue string (GDATLIKYNPADTEVNIALA) containing the mini-loop was selected from 2NPX to substitute for 14 residues (GDCAEYSGIIAGTA) in FLXH by selecting all residues before the first edit point from the first copy of FLXH, the insert string from 2NPX, and the remainder from the second copy of FLXH. A 'merged' model was generated, inspected to verify the edit, and saved as "FXNX.pdb".

The alignment script was modified to reflect the changed template sequence and to realign the ecCx α_7 helix. Finally the python script was modified to specify the start and end residue positions of the helix by deriving a new class, `ecmodel`, from `automodel`, adding the restraint parameters to this before creating the `ecmodel` instance and `make`ing the model.

[1] Position 6 in 2NPX, but 3D structure comparison shows that this is an annotation artefact.

5.5.5.2 Other refinements

Helix α_4 was extended in accordance with the jpred prediction for all NDH2s (P144 to L157), and α_5 was similarly extended (A180 to K202) and split into two separate helices - the consensus for NDH2 predictions. Comparison with the template revealed loci where beta sheet structures had been either curtailed or lost: these were corrected by minor alterations to the alignment file which had been composed manually. A loop between β_5 and β_6 was inserted from 1BHY, the only structure with a similar insertion at this locus. Inspection of the models shows that a glutamine residue (Q332), conserved across all NDH2s and not present in other oxido-reductases is located near the proposed quinone binding site. Consequently the α_7 helix was advanced by two residue positions as this locates Q332 at a position where potential H-bonding with quinone would be possible. (The implications of this are discussed in 5.5.)

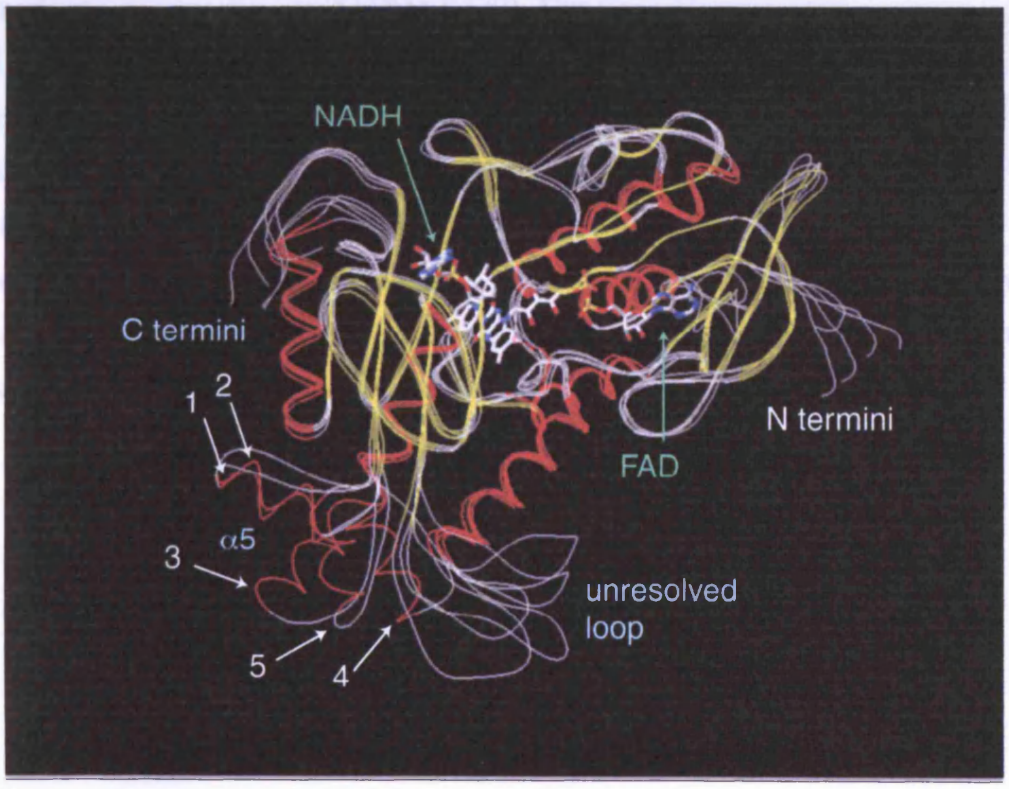


Fig 5.5 Five models generated from the FXNB template illustrating unresolved regions around α_5 and the F158-K170 loop.

Five models were constructed with the final alignment and script configurations and inspected in Spdbv (fig 5.5). In one of these (number 3) the α_5 helix pair is almost coaxial, as would be the case if this was indeed a single helix, but the constraints of the adjacent loop (as modelled) preclude a true full length helix. Model 2 was finally selected partly because of a low reported restraint violation score[1], but also to avoid the implication of a single helix conformation that is not explicitly suggested by the secondary structure prediction.

5.4.5.3 Loop Refinement

One loop region at the exterior of the enzyme (fig 5.6) remained unresolved using the 'chimeric template' technique (F158 to K170). This was addressed using MODELLER's `loopmodel` class from which ten possible solutions were generated. Loop model 2 was selected as this acquires additional stability from inter-residue H-bond interactions, though the decision remains largely subjective. It should be noted that this loop is adjacent to the α_5 region (above, also uncertain) so exhaustive modelling may reveal favourable stabilising interactions. However as it lies at the lower side of the enzyme it is quite likely to interact *in vivo* with the C-terminal domain (not modelled) precluding a definitive conformation.

[1] "sum of all" restraint violations in `model.log` file.

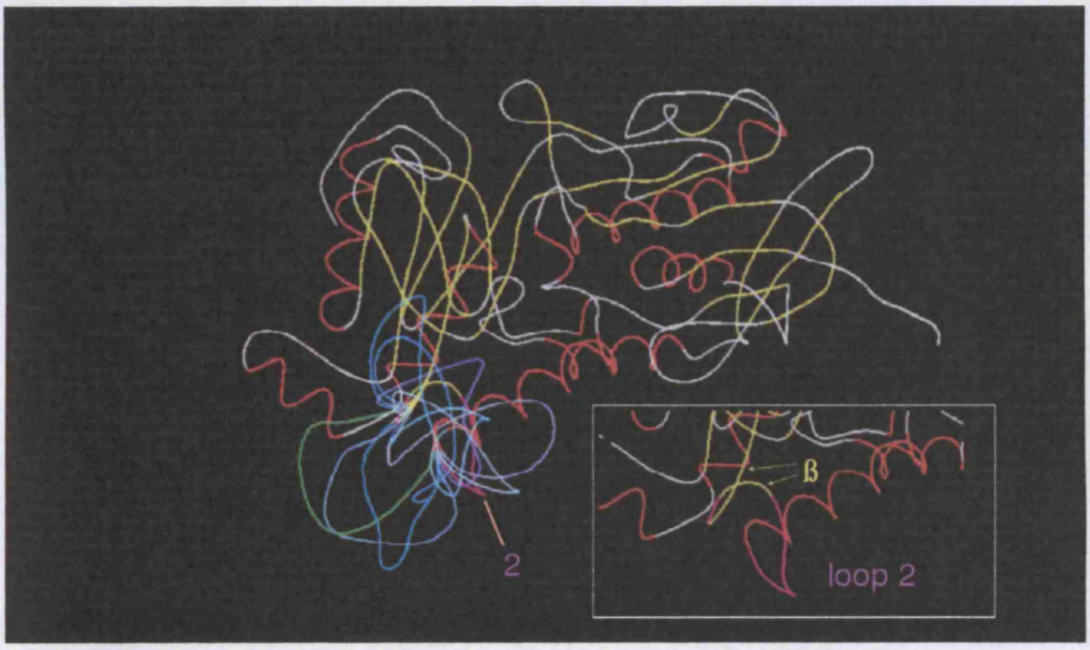


Fig 5.6 Ten models for the F158 - K170 loop

MODELLER was directed to construct ten possible loop regions with the constraints provided (ten different colours in the main diagram).

Inset: model 2 is stabilised by anti-parallel beta strands (marked "β" in the inset). Red: alpha helces; yellow: beta strands; white: loop regions.

5. 5 Analysis of the quinone binding site.

A proposed interaction for ubiquinone with the *E. coli* NDH2 is by 'π-stacking' in a sandwich arrangement with FAD and NADH (Schmid & Gerloff, 2004); this is consistent with known structure for a duroquinone-NAD(P)H interaction (1DXO: Human Nad(P)H-Quinone Oxidoreductase).

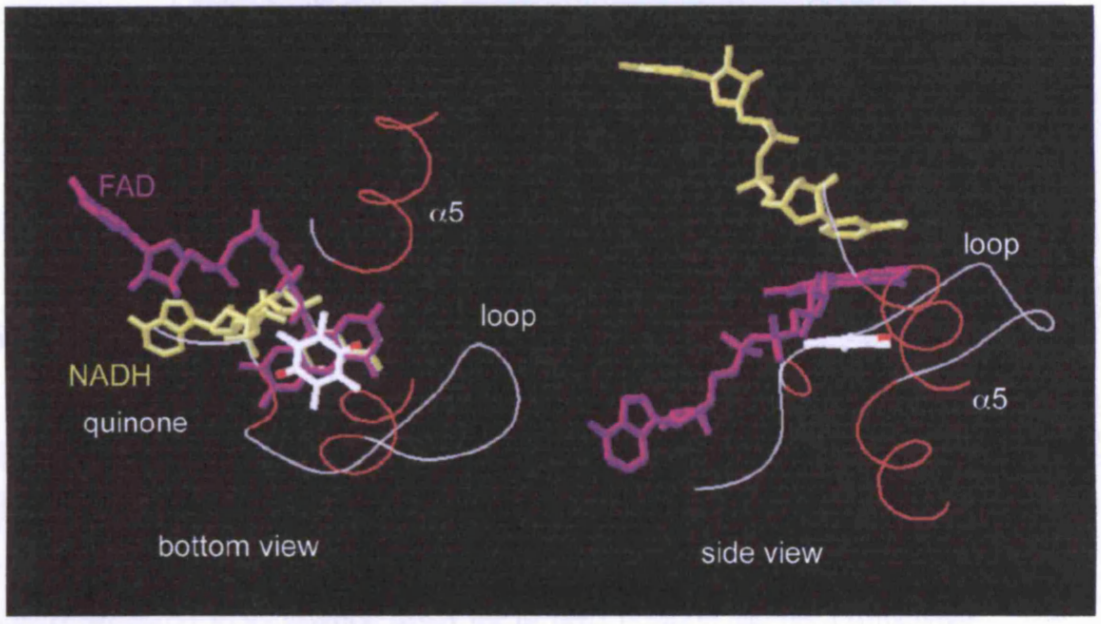


Fig 5.7 Quinone Binding site

View of stacking interactions between NADH → FAD → quinone from two angles. Also shown are peptide regions putatively associated with quinone interaction.

The present model reveals two regions of the protein as being located adjacent to the proposed quinone, and thus offering putative H-bonding residues for interaction with the quinone C=O moieties: the start of α_5 and the H45-D64 loop (fig 5.7). As noted above, Q332 is conserved across all NDH2s suggesting a role as hydrogen donor to the quinone. The loop region contains a number of potential H-bond donors though none (using the current alignment file) is ideally located for a quinone binding role. However, this does not invalidate loop region residues as potential hydrogen donors as this locus is quite divergent from the model (both primary and secondary structure) so its 3D location is necessarily approximate. The most suitable second donor is H45, located at the opposite side of the quinone though this does not result in an entirely convincing topology for quinone binding (fig 5.8).

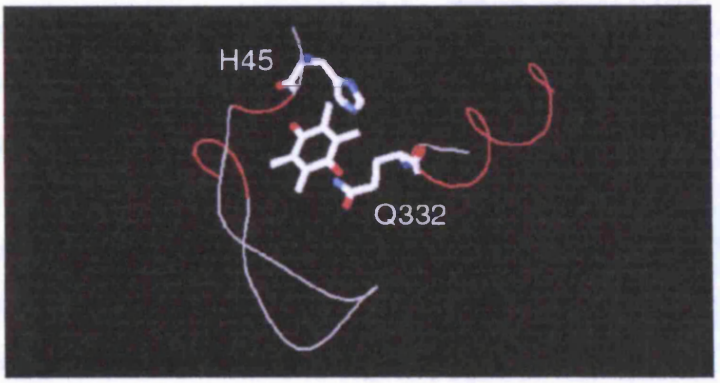


Fig 5.8 H45-Q332 Quinone H-bonding residues

Residues forming proposed hydrogen bonds to the quinone subsrtrate

5.6 An alternative quinone interaction model

Exhaustive attempts at remodelling the *E. coli* NDH2, using alternative template elements and sequence alignments at these two loci, failed to achieve an entirely satisfactory 3D structure at the quinone site. Two other aspects of the model raised further doubts as to its validity: firstly the α_5 helix position had been somewhat contrived in order to position Q332 for a quinone-binding role, and secondly H45 is not conserved in other NDH2s (though alternative adjacent residues may act as potential hydrogen donors).

The assumption of a 'π-stacking' interaction with ubiquinone was made initially because sequence comparison with known FAD/NADH oxidoreductase structures failed to identify similarities (disulphide redox site, etc), and because this topology was found in another structure (1DXO).

Analysis of the NDH2s alignment for the β_2 - α_2 loop region (fig 5.9) failed to find a single-residue locus, both capable of hydrogen donation *and* consistent with the *E. coli* model for a quinone-binding role. However, alignment of a conserved motif ("PLL", *E. coli* P49) reveals a conserved Trp or Phe two residue positions before the motif for most NDH2 species (fig 5.1). Minor realignment of this loop region now located an aromatic

ring adjacent to, and about 3Å from, the FAD ring suggesting a possible alternative electron pathway via this residue to the quinone (fig 5.10).

The α_5 helix was returned to its original conformation, restoring Q336 to its conserved role in stabilising the α_7 - β_{21} loop; this places Q332 in a position to form an H-bond with D60 (also conserved across NDH2s) thus stabilising the β_2 - α_2 loop.

Also, in this model the ubiquinone is not required to migrate as far out of the lipid layer to interact with a redox site.

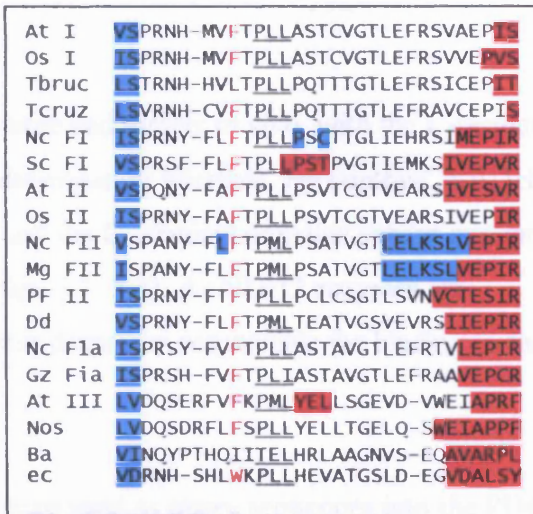


Fig 5.9 NDH2: β_2 to α_2 region

PLL motif underlined, aromatic residues shown in red. (Alpha helix residues shaded red, beta strands in blue).

For these reasons this new model was considered the more plausible topology.

It was not however possible to identify a suitably located, NDH2-conserved, pair of hydrogen donating residues within this structure to be able to propose a putative quinone-binding site.

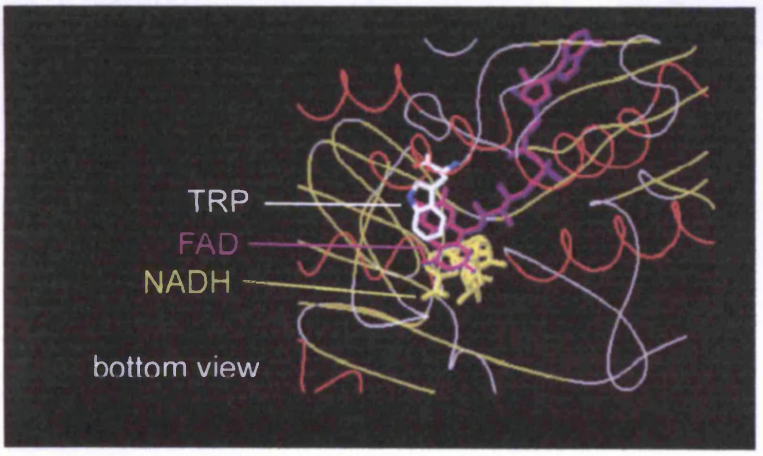


Fig 5.10 Electron pathway in the alternative model

Enzyme viewed from below showing the tryptophan residue proximal to the FAD isoalloxazine ring.

5.7 NDH2 C-terminal domain

Most FAD/NADH oxido-reductases are soluble and dimers *in vivo*, with the C-terminal domains of these enzymes constituting the dimerisation interface. By contrast, NADH2s examined to date are monomeric enzymes, and the C-terminal domains exhibit minimal primary or secondary structure homology (figs 5.1, 5.3). As NDH2 enzymes are membrane bound it has been assumed that this domain is responsible for binding to the lipid bilayer (Schmid& Gerloff, 2004).

The C-terminal domains of the 18 NDH2s were used as query sequences into the PDB database but no homologous structures could be detected, and thus it has not been possible to model this domain.

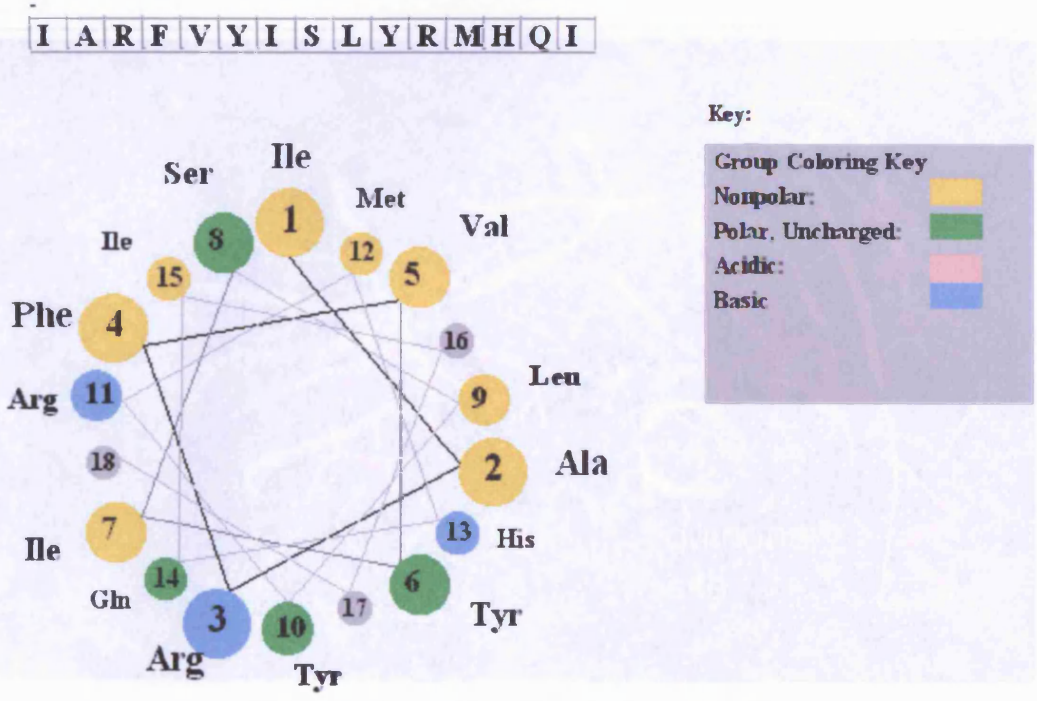


Fig 5.11 Helix wheel representation the *E. coli* C-terminal helix

From examination of Fig 5.1 it may be speculated that the domain comprises a four-strand β-sheet followed by a pair of helices but the secondary structure homology here is weak, other than for the first helix. Helix wheel representations for both putative helices were procured[1] for all 18 NDH2s confirming that all exhibit amphipathic characteristics (Schmid & Gerloff, 2004) (*E. coli* first helix shown in fig 5.11).

5.8 Modelling the *Arabidopsis* NDH2 enzymes

The hybrid template refined for the *E. coli* NDH2 was used for the creation of *Arabidopsis* **nda1** and **ndc** models by manual alignment of their sequences against ecCx and hence FXNB. No anomalous artefacts were detected in these, but the close structural homology to the *E. coli* model is as much a reflection of the modelling process as it is of homologous enzymes (fig 5.12, and Supplementary Data).

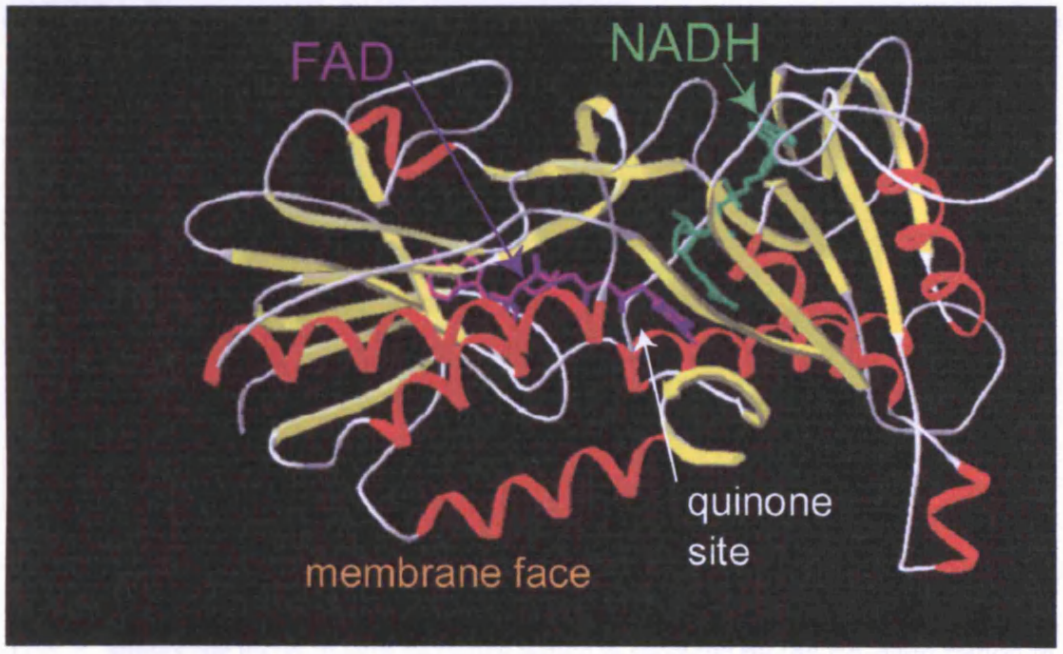


Fig 5.12 The model of *Arabidopsis* **nda1**

Alpha helices are depicted in red, beta strands in yellow. 'Stick' models represent NADH (above) and FAD (centre). The membrane binding region is at the lower face of the enzyme in this orientation.

[1] http://cti.itc.virginia.edu/~cmg/Demo/wheel/wheelApp.html

Each of the four *Arabidopsis* **ndb** enzymes incorporates a pair of contiguous EF-hand motifs (see Chapter 3), inserted between β_{20} and α_7. All NDH2 types have a loop at this locus which has been modelled as in 2NPX (and other known structures), stabilised by a pair of β-strands. The PDB database was searched again (textural searching) in the hope of locating a structure that contains both EF-hand and FAD binding domains, but this was unsuccessful. Models of EF-hands were inspected (1JWD, 1KQV, 1KSM, 1OPQ, 1S6I and 1SL8) and attempts made to introduce manually an EF-hand pair into a modified model of **ndb1** (with this region deleted) but no obvious insertion locus was identified. (A model created in this way would in any case have been highly speculative.)

However, examination of the model shows that the domain would be adjacent to the quinone pocket suggesting that conformation changes due calcium binding would affect accessibility to the pocket, particularly as the C-terminus of α_7 contains "Conserved Region 3", identified above as a putative quinone binding motif.

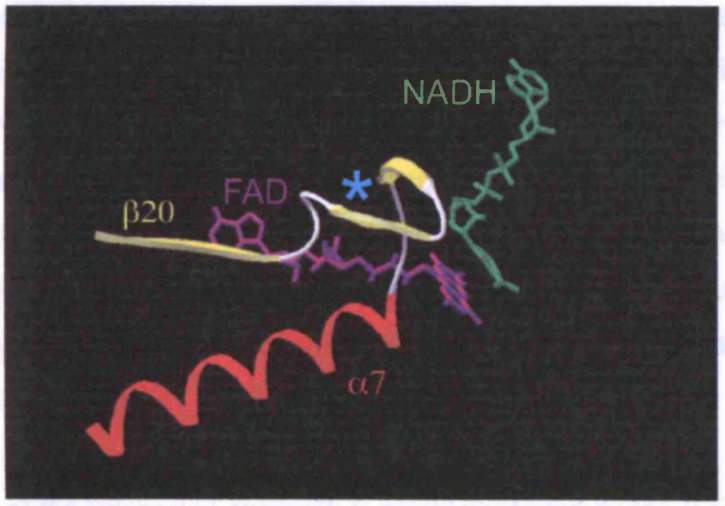


Fig 5.13 The EF-hand locus in **ndb** enzymes

The '*' shows the locus in **nda** enzymes which is replaced by the EF-hand region in ndb NDH2s, between β_{20} and α_7. Distortion of the α_7 helix is thus predicted to influence quinone binding.

Chapter 6. Discussion

6.1 *In silico* analysis.

6.1.1 Modelling technique.

Many proteins may be capable of assuming the mature topology directed only by their amino acid sequence, but state-of-the-art bioinformatics tools (*ab initio* simulation) have yet to achieve this level of automation. Homology modelling, though, has proved a useful structural predictor where a suitable template (experimentally resolved structure) exists.

In the conventional application of this technique a search engine (typically web-based) identifies from a structural database (eg. Brookhaven PDB) the template that exhibits the highest primary and secondary structural homology to that of the target. The modelling software then aligns the target and template sequences and from this creates model structures. These models adopt, naturally, conformations closely matching those of their templates. For example, when the 1OZK 3D structure (Schmid & Gerloff, 2004) is compared visually against those of various resolved FAD/NADH oxido-reductases, its similarity to 2NPX is striking even though its amino acid sequence exhibits little bias towards this enzyme above other family members. A similar situation obtains for the models created in the present work: these have inherited, broadly, the topology of their (in this case, synthetic) template. But while the totality of the models must be viewed with some scepticism, *in silico* 3D analysis provides valuable insights into functional aspects of conserved motifs.

The limitation of the automated approach is amply illustrated by the 1OZK model at the quinone-binding locus (6.1.3). It arises because the homology analysis software cannot distinguish functionally significant loci from the overall sequence. Thus 2NPX (NADH peroxidase), which exhibits the closest homology *overall* was selected as the template, although manual inspection (Chapter 5) reveals that 1XHC (Nitrite reductase) is considerably more homologous *in this particular region.*

A central aim of this research was to develop a more analytic approach to homology modelling. The successful technique proved to be an amalgamation of sequence alignment, 'informed' visual inspection and modelling based on a hybrid (synthetic) template. The broader investigation described here yielded significant benefits and may be summarised as:

1. Compilation and alignment of diverse target (NDH2) sequences; prediction of secondary structure; identification from this of conserved motifs/residues.
2. Compilation of diverse structural family member (FAD/NADH oxido-reductases); extraction of secondary structure information; sequence alignment.
3. Identification of conserved motifs in structural family members.
4. Identification and analysis of divergent features (ie. substrate binding sites).
5. Selection of template polypeptide elements both by sequence homology *and* by considering functional implications.
6. Modelling; analysis and iterative revision of model.

6.1.2 NDH2 is not a disulphide oxidoreductase.

The 3D analysis of second substrate sites in FAD/NADH oxidoreductase family groups has been particularly illuminating. The model analysis show conclusively that NDH2s do *not* incorporate an active disulphide redox site and are thus not "NADH disulphide oxidoreductase" enzymes, as frequently annotated in databases. Furthermore, no cysteine residues pairs have been identified in positions which would permit disulphide bridge stabilisation of tertiary structure, information of value to protein expression and refolding experiments.

6.1.3 Quinone interaction.

Unfortunately it has not been possible to determine a definitive ubiquinone interaction mechanism from the models. The postulation that it involves a direct 'stacking' interaction with FAD (Schmid & Gerloff, 2004) is possible, but the arrival at this conclusion is questionable as it incorrectly presumes a similar topology in 2NPX (NADH peroxidase), conflicting with both the model (steric hindrance by C42) and the latter's authors' analysis (Yeh *et al.*, 1996) and subsequently for 1JOA (Crane *et al.*, 2000). The models proposed here expose the FAD isoalloxazine ring simply because

NADH2s do not contain motifs (disulphide centres, peroxidase C42, etc) that justify the use of such templates. Reassuringly this is consistent with the second substrate-accessible conformation found in the Alkyl hydroperoxide reductase subfamily. Inescapably this represents something of a 'self-fulfilling prophesy'; there may prove to be a novel NDH2-unique interaction, but experimentation with templates and alignments did not elicit an unambiguous solution.

6.1.4 C-terminal domain.

A structure for the C-terminal domain cannot be determined by homology modelling at this time (because of the poor sequence homology to known structures). A review of the secondary structure prediction (fig 5.1 final line) suggests three distinct elements in this domain (ie. following α_7 of the resolved structures, fig 5.3): a putative four strand antiparallel β-sheet, a helix, and either a second helix or pair of β-strands. Of these the first helix is conserved across all members and is shown to be amphipathic (5.7). This and a second amphipathic helix have been proposed as constituting a membrane binding domain (positively charged residues with the hydrophobic face binding phosphatidyl elements: Schmid & Gerloff, 2004). As the predicted secondary structure of the second helix is uncertain (and altogether absent in bacteria Type 2 species) this argument is less convincing. An *E. coli* 'RK loop' (R32-K37) has similarly been proposed as a membrane interacting motif, but this motif is present only in bacteria (and containing basic residues in only a few of Type 1, at that) so this is also speculative.
It is noteworthy that there is considerable residue conservation within the first helix and if, as is expected, this is situated at the membrane-binding face of he enzyme, this may provide the missing elements of the quinone binding pocket.

6.1.5 NADH/NADPH specificity.

In the course of analysing the resolved oxidoreductase structures in order to identify residues participating in H-bond interactions with NAD(P)H (5.4.4, table 5.3), note was also made of those residues participating *exclusively* with NADPH. Is it possible to extract from this data a rule for identifying enzymes which are specific for NADPH?

The pink 'P' region in this table identifies NADPH phosphate group binding residues, from which it can be seen that arginine is the predominant residue, and residue 216[1] (alanine in *E. coli* NDH2) is the most significant locus. How does this correlate with NDH2 enzymes whose substrate has been determined experimentally to be NADPH preferentially?

The *N. crassa* NDE1 enzyme is known to be specific for NADPH over NADH. Its sequence (EAA32649) was examined at this locus but the equivalent residue to *E. coli* A216 was not an arginine. However, there *is* an arginine residue at the equivalent of *E. coli* P222, the second conserved position in NADPH-specific enzymes of the FAD/NADH oxidoreductase family (see table 5.3). The potato NDB sequence was then similarly examined. It too was found to have an arginine at this position.

An alignment was then made using these and the "Representative NDH2" group (Chapter 5). This revealed an arginine at this locus in many enzymes (fig 6.1). These are predominantly the EF-hand-containing enzymes, but also include "Sc FI" which the NDE1 of *S. cerevisiae* (Fungal Type I).

This provides compelling evidence of a predictor for NADH/NADPH specificity from sequence data alone, derived form analysis of the models of related FAD/NADH oxidoreductase family member structures. Not only does this identify a single-residue determinant for NDH2 NADH/NADPH specificity, it also predicts that EF-hand containing enzymes (Fungal Type 2 and Higher Plant **ndb**) are specific for NADPH.

[1] in the *E. coli* sequence

```
nc NDE1 DLTLHFPRLLRNEISVHLIQSRDHILNTYDEAVSKYAEDRFSRDQVDVLVNSRVAEVRPESILF
st ndb1 DLVKIYP-SVKDFVKITVIQSGDHILNTFDERISSFAEQKFQRDGIEVSTGCRVTSVSDHFINM

At I    DVRQRYSH-VKDDIRVTLIEARD-ILSSFDDRLRHYAIKQLNKSGVKLVRGIVKEVKPQKLIL-
Os I    DVRERYAH-VKDYVKVTLIEANE-ILSSFDVGLRQYATDHLSKYGVNLVRGVVKEVKPREIEL-
Tbruc   DVKNINPE-LVQFCKVTVLEAGE-VFSTFDLRVREWGKRRLDALGVRIVKGNVVAVQEKEVIT-
Tcruz   DVRKINHK-LVEFCKVTVLEAGE-VFGMFDLRVRNWGKRRLDALGVRIVKGAVVAVNNKEVVT-
Nc FI   DIKKLIPD-IADRFRVTLIEALPNVLPSFSKQLIEYTESTFKEEKIDIMTKTMVKRVTEKTVEA
Sc FI   DLRKWMPD-LSKEMKVILIEALPNILNMFDKTLIKYAEDLFARDEIDLQVNTAVKVVEPTYIRT
At II   DITKIYPS-VKELVKITLIQSGDHILNTFDERISSFAEQKFTRDGIDVQTGMRVMSVTDKDITV
Os II   DLVKLYPA-IQDFVKITIIQSGEHILNMFDQRIATFAEMKFQRDGIEVNTGFRVVKVSDDLITM
Nc FII  DLTLHFPRLLRNEISVHLIQSRDHILNTYDEAVSKYAEDRFSRDQVDVLVNSRVAEVRPESILF
Mg FII  DLTLHFPKLLRNEISVHVIQSRGHILNTYDETVSKYAEERFARDQVDVLTNSRVSEVKKDRIIF
Pf II   EVKINYKD-IFNFISISIIEGGNNLLPTFTQNISDFTKENFHNLNINVLTNYYVIDVDKHSFHI
Dd      DLSKMFPH-LSKYPKITLVQSADHLLNTFDLKISNYTEKQFERIGIEVLTNTRAVEVKKDHLVV
Nc Fia  DLSKMYPDL-LKFVRITVYDVSPKVLPMFDQALSKYAMDAFKRQKIEIRTQHNIERVRPADGKL
Gz Fia  DLSKLYPDL-MPHVAITIYDIAPKVLPMFDQNLAAYATNIFKREGIRIKTEHHLQGIRRQGDVL
At III  ----------DRGIVQSINVSKNILTSAPDGNREAAMKVLTSRKVQLLLGYLVQSIKRASNLE
Nos     ----------ERGRFRLVEISDQILRTSPDFNREAAKKALDAKGVFIDLETKVESIGQNTISL
Ba      LAKSHGVNP--KEVKLLLVEAGPKILPVLPDHLIERATTSLEARGVTFLTGLPVTNVAGNEIDL
Ec      LHSYGYKGLTNEALNVTLVEAGERILPALPPRISAAAHNELTKLGVRVLTQTMVTSADEGGLHT
```

Fig 6.1 Alignments at putative NADPH determining site

Comparison of known NADPH-specific enzymes (nc NDE1, st **ndb1**) with "Representative NDH2s" at putative substrate specificity locus.
The putative NADPH-determining residues are coloured RED
nc NDE1: *N. crassa* NDE1 (EAA32649); st **ndb1:** *S. tuberosum* **ndb1** (STU245862)
For accession numbers of 'Representative NDH2s" see fig 5.1

6.2 Quinone binding region

6.2.1 Conserved Region 3

An important discovery from the modelling exercise was that "Conserved Region 3", originally identified from sequence data alone as specific to NDH2s, is located at the quinone docking side of the FAD isoalloxazine ring. It may be concluded that this is indeed a quinone-binding region. From examination of this region in species from all groups (fig 6.2) it is clear that a central '**AQ**' is common to all but a few examples (the exceptions are marked '*'). The exceptions are found in Bacteria Type 2, Cyanobacteria, and the Trypanosomes.

These exceptions differ in other respects as well. *Nostoc/Anabaena* species, exceptionally, have no less than six putative NDH2 genes (Chapter 3), yet four of these are of the 'exception' type. If these were excluded then the remaining two would be more consistent with the NDH2 complement of related species. Also a number of the Bacteria Type 2 exceptions have substantial C-terminal extensions (up to 350 additional residues) when compared to 'canonical' NDH2s

The data suggests that the '**AQ**' residue pair may be a determinant for NDH2s and that the exceptions are not NDH2s but a closely related (and novel) enzyme.

An apparently insurmountable impediment to this theory, however, is that it is contradicted by the experimental evidence: the *Trypanosoma brucei* NDH2 is one of these exceptions and it has been characterised (Fang & Beattie, 2003). This would be sufficient to dismiss the theory — unless the *T. brucei* enzyme has been misidentified and is not an NDH2.

Bact Type 1
ec VPPRAQAAHQMAT
vc VPPRAQAAHQMAS
hd IPPRAQAAHQMAT
ms VPPRGQAANQMAT
hi VPPRAQAAHQMAK
wg VPPRAQAAHQMAR
hc VPPRAQSAHQMAS
il VPPRAQSAHQMAE
mc VPPRAQSAHQQAS
av VPPRAQAAHQQAS

Bact Type 2
ll2 TPQIVEAAEQTAH *
li2 TPQIVEGAEQTAL *
bs3 YPPTAQIAMQQGI
dr VPTTAQHAGQQGR
bt HPQLAQVAIQQGE
ba2 HAPSAQLAEGQGE
hs APPTAQAAWQAAW
cs FPPTAQLAEAQGE
st IPQSAQVASQAGS

Cyanobact
av1 VPATAQAAFQQAD
se2 QPGLAQVAYQQGA
pm WPPTAQVALQQGE
ss2 KPALAQIAYQQGA
ns2 QPPTAQVAYQQGI
av2 LPPTAQVAYQQGA
cg LPGVAQVAIQSGE
av6 LSGVAPEALQQGV *
ns6 LSGVAPEALQQGV *

Fungi Type 1
mg1 YGPTAQVASQEGA
sp1 YAPTAQVASQQGA
sp2 LPATAQVANQQGA
sce2FFPTAQVAHQEGE

Fungi Type 1a
mg2 LPKTGQVASQQAV
nci1LPKTAQVAAQQAT
mg2 PPATAQATFQEAK

Fungi Type 2
kl3 LPATAQRAHQQGK
yl2 LPATAQRANQQGV
cn2 YPATAQVASQQGK
um3 LPATAQVAAQQGH

Plant nda
st LPALAQVAERQGK
ps LPALAQVAERQGK
le LPALAQVAERQGK
tb LPTLAAVASRQGV *

Plant ndb
le LPATAQVAAQQGT
zm1 LPATAQVASQQGQ
zm1 LPATAQVAAQEGS
pf PTPTAQNAKQEAY

Plant ndc
at LPTTAQVAFQEAD
os LPATAQVAFQQAD

Fig 6.2 "Conserved Region 3" from a selection of species

Data taken from Figures in Chapter 3. For species identification refer to figs 3.3 to 3.7.
* indicates non-canonical sequence (not glutamine see text)

6.2.2 A critical reappraisal of the *Trypanosoma brucei* NDH2

A significant contribution to NDH2 research has been the successful heterologous expression and characterisation of the *Trypanosoma brucei* putative NDH2 (Fang & Beattie, 2003) following their isolation of NDH2 activity from the inner face of mitochondrial preparations (Fang & Beattie, 2002a). *Uniquely,* the expressed gene proved soluble, stable and exhibited oxidoreductase activity. However, there are compelling reasons to question whether the expressed enzyme is indeed an NDH2.

Firstly it is curious that this NDH2 was successfully expressed as a soluble, stable, active enzyme without any additional manipulation (such as the appending of a soluble domain employed in the current work). This contrasts significantly with the case for the *E. coli* NDH2 (Björklöf *et al.*, 2000) in which the expressed enzyme bound to the membrane, was highly dependent on its environment for activity (detergent composition etc), *and* required particular care in storage conditions (high ionic concentration) to avoid loss of function. Secondly, apart from the discrepancy at "Conserved Region 3", there is a significant difference from other NDH2s in the primary structure: an additional C-terminal region of some fifty residues not found in other NDH2s[1]. (The C-terminus is generally assumed to be associated with membrane-binding.) This cannot be dismissed as a sequencing anomaly as similar extensions are evident in *Leishmania major* and *Trypanosoma cruzi* (see S.5.1), nor as the accidental inclusion of an intron as there are no putative splice sites in this region. Thirdly, and most significantly, the gene encodes *no* mitochondrial targeting signal. Again, this cannot be dismissed as a sequencing anomaly (the erroneous introduction of a start codon, for example) as the same situation exists for *L. major* and *T. cruzi.* It is, at the very least, highly unlikely that this NDH2 - uniquely - has acquired a method for mitochondrial importation that circumvents established mechanisms.

Nor, in fact, is any experimental evidence presented in the paper to link the (previously) detected NDH2 activity to the expressed gene; the latter was selected by interrogation of the emerging *T. brucei* genome database using a consensus NDH2 query string, and it was assumed that the resulting gene must hence also be an NDH2.

A more plausible - if unexpected - interpretation of their data is that this is a (soluble) cytosolic NADH dehydrogenase whose second substrate is 'quinone-like'. This may be the first example of a novel "FAD-linked reductase" enzyme family which has evolved from (or possibly into) the NDH2 family - but which cannot be termed 'NDH2' as it is not coupled to the respiratory chain.

[1] other than more 'exceptions' in the Bacterial Type 2 group

6.2.3. Sequence selection and an NDH2 quinone binding motif

With the *T. brucei* gene now eliminated from the NDH2 family it is necessary to identify other putative 'non-NDH2s'.

Other than the three protozoans mentioned above, all eukaryotic NDH2s identified in this work incorporate the '**AQ**' pair in "Conserved Region 3" and have N-terminal targeting signals. 'Consensus Region 3' is also found in all Type 1 bacterial NDH2s.

However, within the Type 2 bacteria and the cyanobacteria groups there is a number of putative NDH2s which also fail the '**AQ**' selection criterion. Interestingly seven of the Type 2 examples also have significant C-terminal extensions (see S.3.2). By segregating putative NDH2s according to Region 3 homology, tables 3.3 and 3.4 may be retabulated (tables 6.1, 6.2).

Interestingly the abnormally high representation of NDH2s in *Nostoc/Anabaena* species is now reduced to just two. In fact *Bacillus* species now emerge as unusual among prokaryotes in that they encode three NDH2 genes while all other species express none, one or two.

Table 6.1 Bacterial Type 2 NDH2s

Bacterial NDH2s other than *E. coli* homologous and cyanobateria, including genome accession numbers and number of putative genes. Genes partitioned according to "Region 3" homology into 'true' NDH2s and the closely related soluble enzymes ("sol NDH2").

Species	accession no.	NDH2s	sol NDH2
Bacillus anthracis	NC_003995	3	
Bacillus cereus	NC_004722	3	
Bacillus halodurans	BA000004	2	
Bacillus licheniformis	CP000002	3	
Bacillus subtilis	NC_000964	3	
Bacteroides fragilis	CR626927	1	
Bacteroides thetaiotaomicron	AE015928	1	
Clostridium thermocellum	NZ_AABG02000072		1
Deinococcus radiodurans	NC_001263	1	
Enterococcus faecalis	AE016957	1	1
Geobacillus kaustophilus	BA000043	2	
Geobacter sulfurreducens	AE017180		1
Heliobacillus mobilis	AY142850	1	
Kuenenia stuttgartiensis	CT573072	1	
Lactobacillus plantarum	AL935254	1	1
Lactococcus lactis	AE006316		2
Listeria innocua	AL596173	1	1
Listeria monocytogenes	AL591983	1	1
Nitrosospira multiformis	CP000103	1	
Oceanobacillus iheyensis	NC_004193	1	
Oenococcus oeni	NZ_AABJ02000001	1	
Staphylococcus aureus	NC_003923	2	
Streptococcus agalactiae	NC_004368	1	
Thermoanaerobacter tengcongensis	NC_003869		1
Cenarchaeum symbiosum (archea)	DQ397575	1	
Haloarcula marismortui (archea)	AY596297	1	
Halobacterium sp. NRC-1 *(archea)*	AAG19331	1	
Natronomonas pharaonis (archea)	CR936257	1	
Sulfolobus acidocaldarius (archea)	CP000077	1	
Sulfolobus solfataricus (archea)	AE006905	1	1
Sulfolobus tokodaii (archea)	NC_003106	1	

Table 6.2 Cyanobacterial NDH2s

Cyanobacterial NDH2s and homologous species. Genes partitioned according to "Region 3" homology into 'true' NDH2s and the closely related soluble enzymes ("sol NDH2").

Species	accession no.	NDH2s	sol NDH2
Agrobacterium tumefaciens	AE009153		1
Anabaena variabilis	CP000117	2	4
Calothrix viguieri	CVPME131		1
Corynebacterium glutamicum	BX927152	1	
Crocosphaera watsonii	NZ_AADV02000002		2
Gloeobacter violaceus	BA000045	2	
Mycobacterium avium	AE016958	2	
Mycobacterium smegmatis	AF038423	1	
Mycobacterium tuberculosis	AE000516	2	
Nocardia farcinica	AP006618	1	1
Nostoc sp. PCC 7120	BA000019	2	4
Prochlorococcus marinus	BX572099	1	
Rhodopirellula baltica	BX294142	1	
Synechococcus elongates PCC 7942	CP000100	2	1
Synechocystis sp. PCC 6803	BA000022	2	1
Solibacter usitatus	NZ_AAIA01000044		1
Trichodesmium erythraeum	NZ_AABK04000002		1

6.3 Sequence analysis

Current automated amino acid sequence and phylogenic analysis programs provide an efficient system for the rapid processing of vast databases. However, they are necessarily unaware of the significance of conserved loci so cannot distinguish between 'functional' and 'coincidental' homology, which can elicit misleading results. This phenomenon is particularly problematic in cases such as NDH2 where a number of similar but distinct family members exists (ie. FAD/NADH-linked oxidoreductases).

In the present species-wide compilation putative NDH2 sequences have been selected on the basis of conservation at particular loci (Chapter 3) then categorised by homology. Thus many previously suggested NDH2s have been omitted (eg. Michalecka *et al.*, 2003: figure 2, including three additional *Nostoc* genes)[1] while others have been included. The results show that bacteria sequenced to date encode generally one NDH2, occasionally two, or none (on the evidence of completed genomes). Species with the shortest N-termini have just three residues prior to the consensus `IVIV` (the first β-strand), suggesting that bacterial NDH2s diffuse to the periplasmic membrane rather than employing targeting signals. The cyanobacteria encode a single NDH2 gene which is significantly divergent from other bacterial groups.

Fungal species (and *D. discoideum*) express two or three NDH2 proteins, probably always including at least one of Type 1. *S. cerevisiae*, which has been subjected to the most extensive biochemical investigation, appears to be atypical in that all NDH2s are of type 1. A shared ancestral development is probably responsible for the emergence of Type 1a proteins in some fungal species (the functional distinction of which from Type 1 cannot be determined); but more interestingly the third NDH2 in some species has acquired a putative EF-hand domain similar to the **ndb**s in higher plants.

Paucity of completed *viridiplantae* genome sequences precludes any generalisation concerning NDH2 complement in these species but the presence of at least three **ndb** genes in *O. sativa* (as well as four in *Arabidopsis*) implies that two gene duplication events must have occurred before the divergence of, or possibly separately within both,

[1] Also in this paper, *N. crassa* NDE1 has been placed with the higher plant NDB group rather than the fungal group, an artefact of automated analysis.

monocotyledons and dicotyledons. A similar duplication event must account for the two **nda** genes. The sequence of the single **ndc** gene is consistent with the model of an ancestral endosymbiotic relationship with a cyanobacterium and, as there is no evidence of a pseudogene in the completed *Arabidopsis* genome, the incorporation of symbiont genes into the host nuclear genome would appear to postdate the **nda** and **ndb** duplication events.

The relationship between NDH2 sequences can be expressed diagrammatically (fig 6.3). An evolutionary pathway may be inferred but this may be misleading: for example, while *viridiplantae* species exhibit close sequence homology between **nda** and **ndb** genes (when compared to fungal Type 1 and 2), it is equally possible that a single EF-hand acquisition event, prior to fungal and higher plant divergence, was followed by convergent evolution of genes within these species groups.

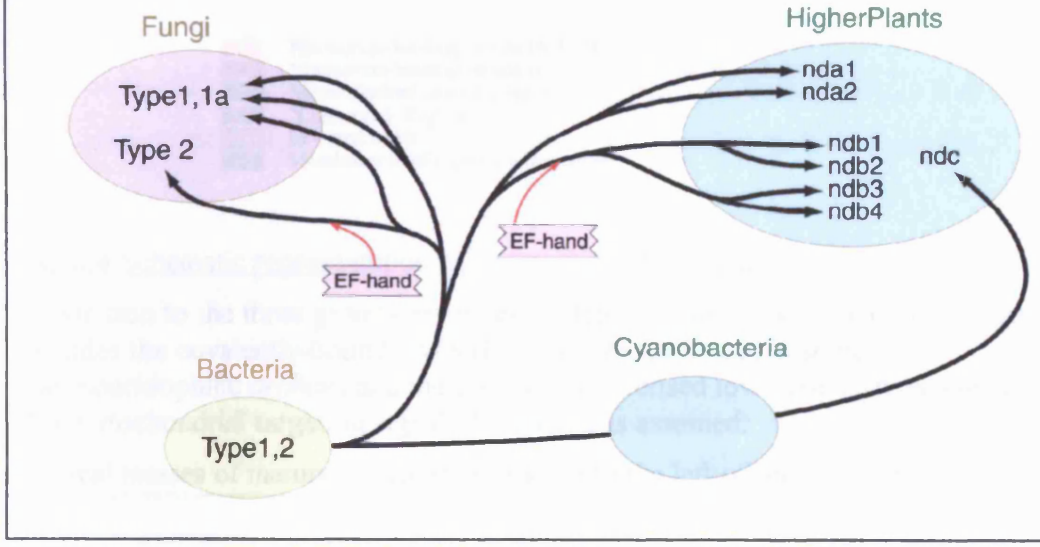


Fig 6.3 NDH2 Relationships amongst species

Grouping is according to homology. Red arrows indicate possible incorporation events of an EF-hand domain.

See Chapter 3 for the full classification of NDH2 types

Finally the various genes which have been demonstrated to exhibit NDH2 activity may be represented schematically (fig 6.4). Five distinct structures may be distinguished: the three evolutionarily related NDH2s examined in detail in this research, the FMN containing gene found in extreme bacteria, and the 33 kDa protein which has been purified from mitochondrial preparations but for which no sequence has been identified to date. The figure illustrates the principle functional elements in these enzymes (the mitchondrial targetting signal for the 33 kDa protein is assumed).

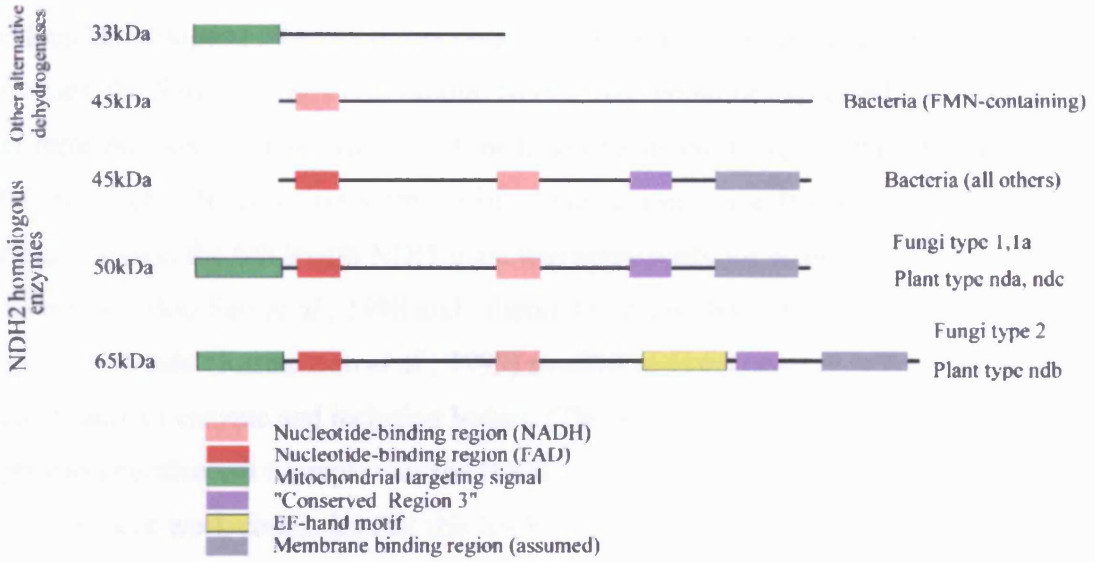


Fig 6.4 Schematic representation of alternative NADH dehydrogenases

In addition to the three groups examined in depth in this work (NDH2s) this figure includes the covalently-bound FMN dehydrogenases (found in some thermoacidophilic *archae*) and the poorly characterised low mass enzyme (see text). The mitochondrial targeting signal of the latter is assumed.

Typical masses of mature enzymes are shown to the left of the schematics.

6.4 *In vitro* analysis

The RNA isolation, mRNA purification, RT-PCR and cDNA cloning experiments demonstrate that *Arabidopsis* NDH2 genes can successfully be obtained. **nda1** was obtained with little difficulty, which is probably attributable to its relative mRNA abundance in leaf tissue (Michalecka *et al.*, 2003). Digest analysis of clones derived from other genes suggests that these too can be obtained though optimum conditions are required: initial mRNA purification (from total RNA) and a 1:1 mix of Sensiscript and Omniscript transcriptase enzymes in the RT-PCR reactions.

Previous heterologous NDH2 expression experiments have resulted in a variety of outcomes: the *Saccharomyces cerevisiae* NDH2 gene could be expressed in *E. coli* only after removal (possibly coincidental) of the targeting signal (Kitajima-Ihara & Yagi, 1998) although activity assays were performed using membrane fractions rather than purified enzyme; the full length NDI1 gene was appropriate for hamster cell transfection. (Boo Seo *et al.*, 1998 and others). Overexpression in *E. coli* of the full-length potato **nda** (Rasmusson *et al.*, 1999) resulted in a combination of both membrane located (active) enzyme and inclusion bodies. (The only 'truly' successful heterologous expression/purification example was the His-tag appended *Trypanosoma brucei* gene but the present work concludes that this has been erroneously identified an NDH2.)

Both pET32 and pET22-based expression systems yielded high levels of NDH2 protein - as much as 50% of total protein for p32-718 as visually determined from SDS-PAGE analysis. However most (probably all) of the expressed protein aggregated as inclusion bodies, as is commonly the situation for exogenous membrane proteins. Attempts to produce an active protein by varying the induced culture growth conditions did not circumvent this problem.

Extensive digest analysis of the clones, using multiple restriction enzymes, produced the predicted fragment pattern. This provides compelling evidence for a correctly cloned cDNA insert, but it is always possible that a point mutation (or other frame-coherent error) has occurred at a critical residue (or residues). Definitive validation could be obtained by full-length DNA sequencing. While FAD was included in the growth medium at various concentrations it is possible that the *E. coli* cell wall is not permeable

to FAD and that the absence of a cofactor in sufficient abundance is responsible for the failure to express active protein, although it is expected that endogenous levels would suffice for detection of trace enzyme activity as was the case for *E. coli*. A third possible cause for the failure to express active protein, also applicable to the failure to refold the inclusion bodies, is the N-terminal region. AT611 incorporates the full-length translated sequence while p32-718 omits the predicted cleaved targeting signal. However, sequence comparison shows that eukaryotic N-terminal regions are highly divergent in length, amino acid composition and predicted secondary structure. It is tempting to speculate that the entire region, prior to the equivalent bacterial initiation residue, is cleaved following mitochondrial import[1]. Whether this is problematic is less clear. The model for the *E. coli* NDH2 reveals that the N-terminus is at the outer face of the enzyme at the predicted soluble (cytosolic) side so an extension here is less likely to impede correct folding. Indeed pET22 appends a modest (31 residue), and pET32 a substantial (~140 residue) N-terminal fusion product, both of which have established track records of successful expression, though this is an enzyme-specific phenomenon. The possibility cannot be eliminated that formation of tertiary structures in this region serves to inhibit enzyme folding prior to export in the native environment, so identification of the mature N-terminus locus would be of benefit. Finally, it cannot be ruled out that a chaparone (or other *Arabidopsis*-native enzyme) is an essential contributor to the assembly of the mature tertiary structure.

The expression experiments show that exogenously expressed *Arabidopsis* NDH2 aggregates either predominantly or exclusively as inclusion bodies under all growth conditions tried. The experiments show also that expression yield is high (up to 50% of total protein). As isolation of inclusion bodies by fractionation alone can result in purity sufficient for crystal growth an alternative expression regime is possible: the gene(s) could be subcloned into an expression vector as 'native' NDH2 (ie. without any fusion elements). This would both remove the need for downstream processing (*enterokinase* cleavage of pET-32 products and subsequent purification), and remove the reservations associated with the presence of N-terminal fusion elements. Further, constructs could be made in which the entire eukaryotic N-terminal region is also excised to exclude the possibility that this is inhibiting the folding process.

[1] p32-718 retains some 40 residues, AT611 the full length of around 80.

The refolding experiments failed to realize evidence of NADH dehydrogenase activity as determined by oxidation using FeCN, an established assay for similar enzymes (Moore *et al.*, 2003; Barber & Quinn, 2001). This assay has also been used with FAD/NADH oxidoreductase enzymes which have a putative tryptophan electron intermediary (eg. ferredoxin reductase: Jose Quiles, 1998) so even if the alternative NDH2 model is applicable (Chapter 5) the FeCN assay should still be valid.

There is no report of previous attempts to refold NDH2 inclusion bodies, though aggregation of heterologous *E. coli* expression of a higher plant gene has been observed (Rasmusson *et al.*, 1999). The sensitivity to environmental conditions of initially active purified NDH2 protein (Björklöf *et al.*, 2000, and others) implies that successful refolding may be critically dependent upon selection of the correct conditions. An infinity of refolding environments exists, one of which may prove applicable to the refolding of p32-718 and/or AT611. However, the modelling task has revealed an absence of potential disulphide bridges so redox poise (eg GSH:GSSG), for this purpose at least, need not be of concern.

An obvious focus of attention for the refolding of membrane-associated proteins is the provision of a detergent/lipid environment. This may be true, but the modelling analysis suggests that this is not likely to explain the failure of the FeCN assays. The domains responsible for FAD binding and the NADH pocket are homologous to those of the soluble enzymes (for which structures have been resolved) while the N-terminal domain probably serves as a membrane interface. This implies that correct folding of the FAD/NADH domains should *not* be dependent upon prior N-terminal domain assembly. (Reaggregation may occur but some NADH-FeCN oxido-reductase should be detectable.)

6.5 Concluding observations

The failure to obtain enzymatically active *Arabidopsis* NDH2 - either through manipulation of the expression environment or by refolding of inclusion bodies - was a disappointment. Curiously, there are reports in the literature of his-tagged expression of NDH2 genes (eg. potato NDA and NDB, Rasmusson *et al*, 1999) which make no mention of protein purification, although this is the principal purpose of appending the tag. Indeed a yellow colour is reported when tagged *Saccharomyces cerevisiae* NDI1 was loaded onto an affinity column (Kitajima-Ihara & Yagi, 1998), yet there is no ensuing report of characterisation of the bound protein. It would be, to say the least, curious that these groups did not attempt to purify their expressed protein. It seems likely that repeated attempts have indeed been made but without success and hence not reported. In fact, the only report of successful heterologous NDH2 expression is that for *Trypanosoma brucei* (Fang & Beattie, 2003), but this has been shown in the present work to have been misidentified.

With the benefit of hindsight the *Arabidopsis* NDH2 was an overly ambitious target protein for heterologous expression. Furthermore, given the environmental constraints required to maintain activity for the over-expressed *E. coli* enzyme (even in the short to medium term) crystal formation will be a particular challenge.

By contrast the computational and sequence analysis has yielded numerous novel insights into NDH2 characteristics:

Analysis of sequence data initially identified "Conserved Region 3" as an NDH2-specific motif, a proposal subsequently confirmed by homology modelling which located this at the quinone binding site of the enzyme. This discovery of this novel quinone-binding motif was hindered by at first by the *T. brucei* sequence which appeared to contradict this theory. A reappraisal of the data published for the latter enzyme determined (for unrelated reasons) that it has been misidentified as an NDH2 and is more probably a (soluble) cytosolic enzyme. The existence of bacterial sequences that similarly differ at this region (but are otherwise highly homologous to NDH2s) suggests the existence of a (novel) FAD/NAD-linked enzyme family, related in an evolutionary sense to the NDH2s but probably not associated with the respiratory chain.

Homology modelling has proved particularly fruitful - most productively when multiple structures are compared. The most elegant example of this is the identification of a residue specific to NADPH (as opposed to NADH) by comparing hydrogen bonding patterns of structures (other than NDH2s) of known substrate specificity. The resulting theory was then tested by applying this to NDH2s of known specificity: this correlated perfectly with the proposed theory. This residue may be considered a substrate specificity marker applicable to all FAD/NAD-linked reductase family members.

By employing chimeric templates for the *E. coli* model it was possible to demonstrate that the existing model (Schmid & Gerloff, 2004) was incorrectly modelled at the crucial quinone-binding region and two plausible alternatives were proposed. One of these assumes a direct interaction ("π-stacking") between the quinone and the FAD ring as previously proposed (but with a model that now permits this interaction). The second, in which electrons pass via an aromatic intermediate, was arrived at by noting conserved aromatic residues in aligned NDH2 sequences (other than type 2 bacteria) at a locus suggested by the model. This amalgamation of structural modelling with sequence alignment has proved a powerful (and novel) technique for the identification of motifs within sequence data.

In spite of the fact that it is not possible to model an EF-hand containing NDH2 (as no appropriate structure has been physically determined) it was possible to identify the regulatory domain insertion point by comparison of *Arabidopsis* **nda1** and **ndb1** sequences. From this it could be concluded that calcium regulation operates by introducing conformational distortion of the quinone binding site and thus regulation as achieved by modulation of the binding affinity to this substrate.

Examination of the structural model has also revealed the absence of proximal cysteine residues in *Arabidopsis* nda1, indicating that future attempts to express this protein need not address the issue of redox state for disulphide bridge formation.

But the most intriguing question about these enzymes remains unanswered: Why, particularly given that their very existence is superfluous to the requirements of mammalian mitochondria, are there so many of them? There must be a good reason for

Arabidopsis to maintain seven discrete mitochondrially-targeted NDH2 genes. Certainly they have evolved a limited degree of functional divergence: the optional EF-hand domain and NADH/NADPH specificity as enzymatic discriminators, and targeting signals which direct them to two different faces of the membrane. However, their cellular raison d'être is less clear - except perhaps in the case of *Arum maculatum* and other species which apply them to the task of thermogenesis. But this does not explain their role in the majority of species. Evidence of elevated expression in the presence of light (for some) provides one piece of the jigsaw, while various regulatory functions may contribute a few more. But as alternative pathways exist for their *enzymatic* functions (redox shuttles, UCPs), their colonisation of genomes is impressive.

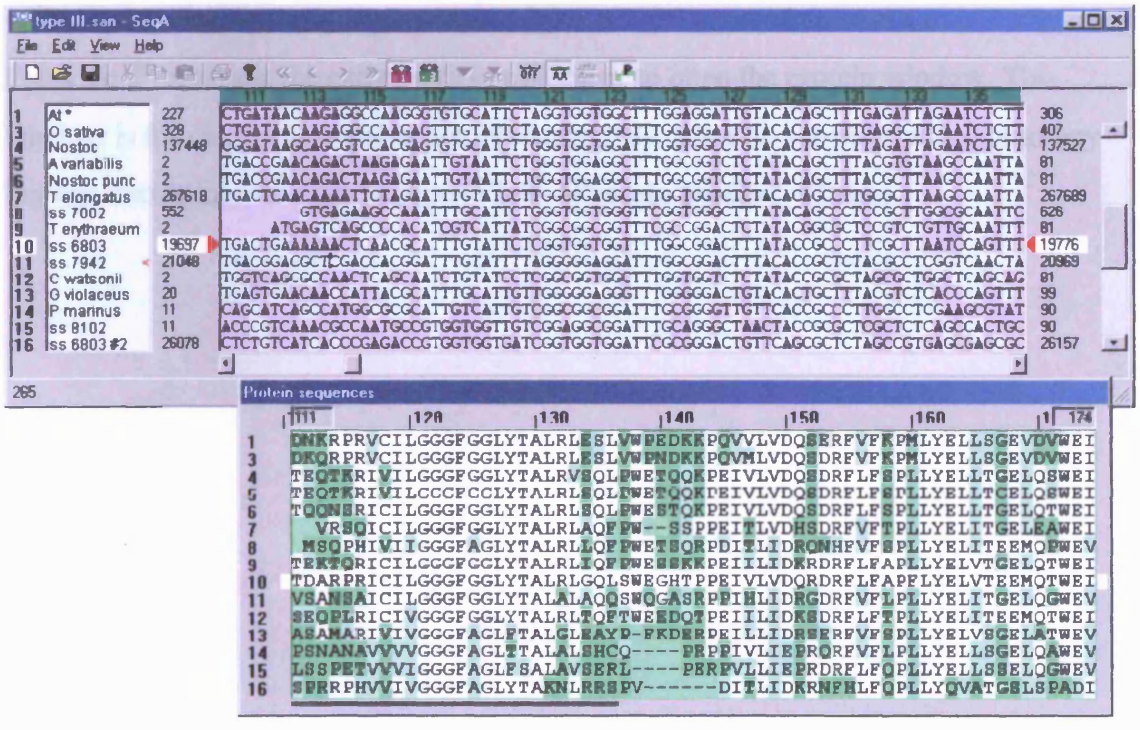
Further elucidation of the functional and mechanistic properties of the alternative NAD(P)H dehydrogenases will provide tangible answers to *how* they perform, but the very breadth of the repertoire (in higher plants, at least) must simply pay testimony to the importance of a exquisitely tuned mechanism for coupling electron transport to oxidative phosphorylation.

Appendix

Software Applications

Microsoft Windows® applications were written to assist in the analysis of bioinformatics data and for the recording and display of spectrophotometer time-course traces. Code for these was written in C++/MFC and PIC assembly language.
These are included on the accompanying Supplementary Data CD and may be used to view the supplementary files. Many features are not pertinent to this task (and not always 'user friendly') so the instructions for use included below are just the minimum required to permit examination of the CD files. Each may be executed directly from the CD or copied to computer; no installation process is required.

A.1 **SeqA** Nucleotide and Protein Sequence Analysis



This program has two windows: one for nucleotide sequences and the other to display the corresponding polypeptide sequences. It was written originally to collect and align sequences and includes various additional features such as pattern finding and homology searching. A new entry is created by copying a nucleotide sequence to the clipboard then selecting **File→Input from Clipboard.** Manual alignment is achieved using the standard editing keys and the spacebar. Advanced functions are accessed by right-clicking in either the main nucleotide view or the name list area (at left).
A particularly useful feature of this application (unique?) is to visualize the relative similarity of one gene to the others in the collection (both by nucleotide and protein residue) by simply clicking on that line. The background colouring changes to reflect the new homology.

Simple instructions for use:

Double-click the icon to start the application.

Drag and drop a sequence collection file (`.san`) onto the window to open the file (or use **File→Open**). For larger sequence collections the window may be enlarged by dragging its lower edge, or the right-hand scroll-bar used to view different entries. The display is coloured to identify differences between the selected sequence and the collection. (Click on a sequence to select it.)

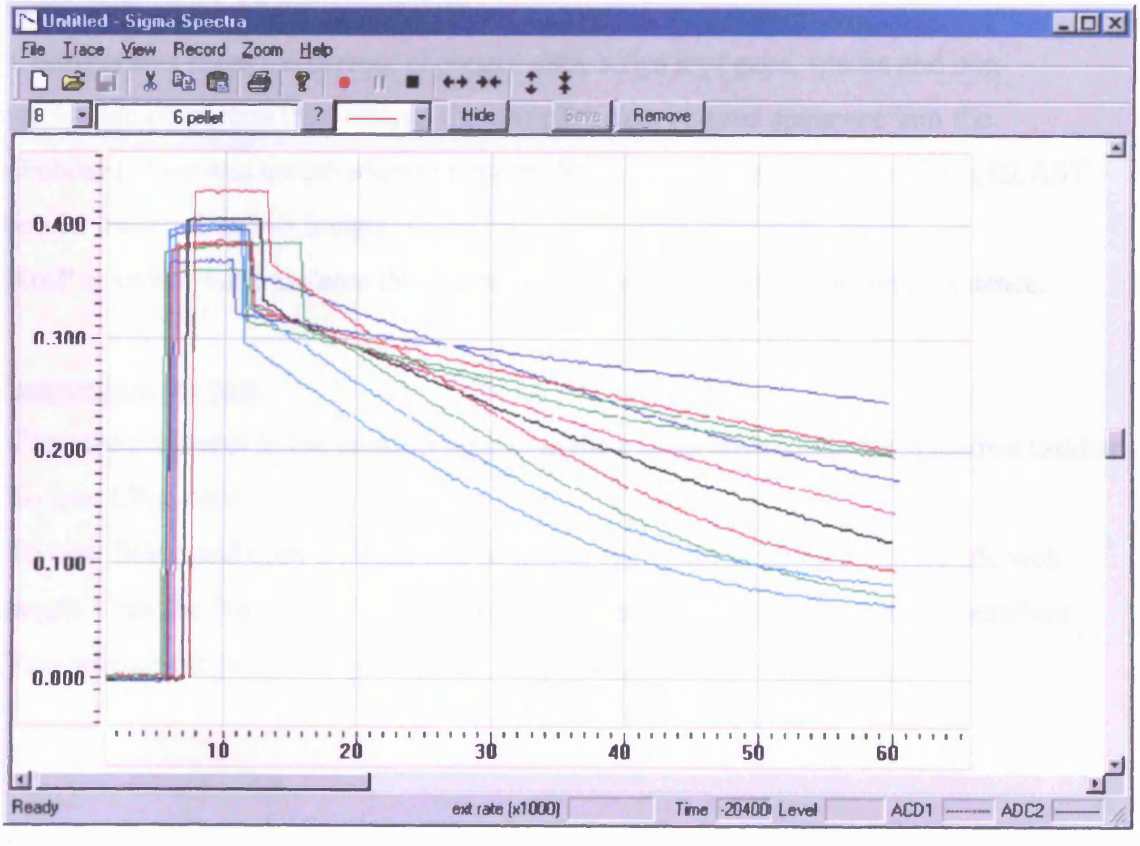
Press the **P** button at the extreme right of toolbar to open the protein window. This window is for display only and is not interactive. By default it shows peptide homology but can also colour residues by type (**View→Protein Display→Type**).

A.2 **Spectra** Spectrophotometer data logger and display

The dual-beam spectrophotometer has an analogue output signal but no computer interface. This application was written, and an external interface unit constructed, so that time-course traces could be recorded directly into a Windows environment and visually assessed.

The external unit employs a 12-bit anolog-to-digital converter (ADC) and a PIC 18F84 microcontroller programmed in assembly language to perform conversions and send RS232 data to the PC Serial port (Com1) at 4Hz. The ADC resolution ensures accuracy is maintained to that of the spectrophotometer (A = 0.001) and the interface calibrated to track the display on the latter over the range A=0.000 to A=1.000. The use of low voltage/low power circuitry allows this to be powered from spare serial port signals obviating the need for a battery or PSU.

The Spectra windows application has Record/Pause/Stop controls for real-time data input and simultaneous display of previous traces.



Simple instructions for use:

Double-click the icon to start the application.

Drag and drop trace files (`.trc`) to add to the display (or use **File→Open**).

Use standard zoom controls and scrollbars to adjust the view window, or drag the mouse over the horizontal or vertical scales.

Click near any trace to select it (or select from the numbered pull-down control at the extreme left of the toolbar). Press '?' to see any text description for this trace.

Extinction change rates ($\Delta A/\Delta t$) may be measured by dragging the mouse between any two points in the display creating a black rectangle. The value is displayed in the status bar (at the bottom of the window.

A.3 **Compstrand** and **NtoP** utilities

These simple utilities operate on clipboard data and saved much time providing "one click" conversion on nucleotide data.

Compstrand takes the current clipboard data, strips it of gaps, spaces and non-nucleotide characters then returns the complementary strand sequence into the clipboard. This was useful when generating PCR reverse primers and copying BLAST results on -1, -2 and -3 frames.

NtoP is similar but translates the nucleotide sequence into an amino acid sequence.

Instructions for use:

Copy the programs to the desktop then drag their icons down onto the Windows taskbar.

To use: Click once.

To test: Select and copy a nucleotide sequence (eg. from a Genbank nucleotide web page). Click the **NtoP** icon. Paste into a text document or BLAST protein query box. You will see the protein sequence.

S. Supplementary Data

This section contains additional bioinformatics data and an index to the files included on the accompanying CD.

Sequence Alignments (CD:→Alignments)

S.1.1 **Arabidopsis →AtNDHs.san**

The nine *Arabidopsis* genes detected by TBLASTN with *nda1* as query string.

S.1.2 **Arabidopsis →At1g07180.san**

Genomic, cds and EST Genbank database submissions for *Arabidopsis nda1* identifying introns and exons. Also contains PCR primers and amplification product.

S.1.3 **Arabidopsis →At2g29990.san**

Genomic, cds and EST Genbank database submissions for *Arabidopsis nda2* identifying introns and exons. Also contains PCR primers and amplification product.

S.1.4 **Arabidopsis →At4g21490.san**

Genomic, cds and EST Genbank database submissions for *Arabidopsis ndb3* identifying introns and exons. Also contains PCR primers and amplification product.

S.1.5 **Arabidopsis →At4g05020.san**

Genomic, cds and EST Genbank database submissions for *Arabidopsis nda2* identifying introns and exons. Also contains PCR primers and amplification product.

S.1.6 **Arabidopsis →At4g28220.san**

Genomic, cds and EST Genbank database submissions for *Arabidopsis ndb1* identifying introns and exons. Also contains PCR primers and amplification product.

S.1.7 **Arabidopsis →At2g28200.san**

Genomic, cds and EST Genbank database submissions for *Arabidopsis ndb4* identifying introns and exons. Also contains PCR primers and amplification product.

S.1.8 **Arabidopsis →At5g08740.san**

Genomic, cds and EST Genbank database submissions for *Arabidopsis ndc1* identifying introns and exons. Also contains PCR primers and amplification product.

S.1.9 **Arabidopsis →At5g22140.san**

Genomic, cds and EST Genbank database submissions for *Arabidopsis* 5g22140 identifying introns and exons. Also contains PCR primers and amplification product.

S.1.10 **Arabidopsis →At3g44190.san**

Genomic, cds and EST Genbank database submissions for *Arabidopsis* 3g44190 identifying introns and exons. Also contains PCR primers and amplification product.

S.1.11 **Arabidopsis →At611.san**

PCR primers and amplification product.

S.2.1 **Oryza →OsNda.san**

Genomic and mature mRNA sequence for the two *Oryza sativa nda* genes identifying introns and exons.

S.2.2 **Oryza →OsNdb.san**

Genomic and mature mRNA sequence for the three *Oryza sativa ndb* genes identifying introns and exons.

S.2.3 **Oryza →OsNdc.san**

Genomic and mature mRNA sequence for the *Oryza sativa ndc* genes identifying introns and exons.

S.3.1 **Bacteria→Bacteria1.san**

Alignment of putative NDH2 genes homologous to *E.coli.*

S.3.2 **Bacteria→Bacteria2.san**

Alignment of other putative bacteria NDH2 genes

S.3.3 **Bacteria→Bacteria3.san**

Alignment of cyano- and related bacteria NDH2 genes

S.3.4 **Bacteria→FMN.san**

Alignment of FMN containing NDH2s in *archea* species

S.4.1 **Fungi→Fungi1.san**

Alignment of putative *Fungi* Type 1 NDH2 genes.

S.4.2 **Fungi→Fungi1a.san**

Alignment of putative *Fungi* Type 1a NDH2 genes.

S.4.3 **Fungi→Fungi2.san**

Alignment of putative *Fungi* Type 2 NDH2 genes.

S.5.1 **Plant→nda.san**

Alignment of *viridiplantae* **nda** (includes *Trypanosoma* genes)

S.5.2 **Plant→ndb.san**

Alignment of *viridiplantae* **ndb** (includes *Plasmodium* genes)

S.5.3 **Plant→nd3.san**

Alignment of *viridiplantae* **ndc** genes

Protein Modelling (CD:→Models)

(PDB files are presented both with and without FAD/NADH ("HETAM") molecules. See also the ReadMe file on the CD)

S.6.1 **ec model1→ model1.pdb**
ec model1→ model1(no HETAM).pdb

The initial *E. coli* NDH2 model

S.6.2 **ec model1→ source** (directory)

Compilation files for the above

S.6.3 **ec model2→ model2.pdb**
ec model2→ model2(no HETAM).pdb

The alternative *E. coli* NDH2 model

S.6.4 **ec model2→ source** (directory)

Compilation files for the above

S.6.5 **at nda1→ nda1.pdb**
at nda1→ nda1 (no HETAM).pdb

The *Arabidopsis* NDH2 model

S.6.6 **at nda1→ source** (directory)

Compilation files for the above

S.6.7 **ec support pdb files** (directory)

Intermediate files used in the development of the *E. coli* templates

S.6.7 **additional files** (directory)

FAD/NADH-linked oxidoreductase (and other) source structure files reviewed in process of developing the structural models.

References

Affourtit C, Krab K & Moore AL (2001) "Control of plant mitochondrial respiration" *Biochim. Biophys. Acta* **1504** 58-69

Altschul, SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, and David J. Lipman DJ (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs" *Nucleic Acids Res.* **25** 3389-3402.

Bakker BM, Bro C, Kötter P, Luttik MAH, van Dijken JP & Pronk JT (2000) "The mitochondrial alcohol dehydrogenase Adh3p is involved in a redox shuttle in *Saccharomyces cerevisiae" J. Bacteriology* **1872(17)** 4730-4737

Bakker BM, Overkamp KM, van Maris AJA, Kötter P, Luttik MAH, van Dijken JP & Pronk JT (2001) "Stoichiometry and compartmentation of NADH metabolism in *Saccharomyces cerevisiae*" *FEMS Microbiol. Rev.* **25** 15-37

Bandeiras TM, Salgueiro C, Kletzin A, Gomes CM & Teixeira M (2002) "*Acidianus ambivalens* type-II NADH dehydrogenas: genetic characterisation and identification of the flavin moiety as FMN" *FEBS Lett.* 531 273-277

Bandeiras TM, Salgueiro CA, Huber H, Gomes CM & Teixeira M (2003) "The respiratory chain of the thermophilic archeon *Sulfolobus metallicus*: studies on the type-II NADH dehydrogenase" *Biochim. Biophys. Acta* **1557** 13-19

Baranova EA, Holt PJ & Sazanov LA (2006) "Projection Structure of the Membrane Domain of *Escherichia coli* Respiratory Complex I at 8 Å Resolution" *J. Mol. Biol.* **366(1)** 140-154

Barber MJ & Quinn GB (2001) "Production of a recombinant hybrid hemoflavoprotein: engineering a functional NADH:cytochrome c reductase" *Protein Expr. Purif.* **23(2)** 348-358

Bertsova YV, Bogachev AV & Skulachev VP (2001) "Noncoupled NADH:ubiquinone oxidoreductase of *Azotobacter vinelandii* is required for diazotrophic growth at high oxygen concentrations" *J. Bacteriology* **183(23)** 6869-6874

Bhattacharya S, Bunick CG & Chazin WJ(2004) "Target selectivity in EF-hand calcium binding proteins" *Biochim. Biophys. Acta* **1742** 69– 79

Björklöf K, Zickermann V & Finel M (2000) "Purification of the 45 kDa, membrane bound NADH dehydrogenase of *Escherichia coli* (NDH-2) and analysis of its interaction with ubiquinone analogues" *FEBS Lett.* **467** 105-110

Bonner WD & Voss DO (1961) "Some characteristics of mitochondria extracted from higher plants" *Nature* **191** 682-684

Boo Seo B, Kitajima-Ihara TK, Chan EKL, Scheffler IE, Matsuno-Yagi A & Yagi T (1998) "Molecular remedy of complex I defects: rotenone-insensitive internal NADH-quinone oxidoreductase of *Saccharomyces cerevisiae* mitochondria restore the NADH oxidase activity of complex I-deficient mammalian cells" *Biochem.* **95(16)** 9167-9171

Boo Seo B, Matsuno-Yagi A & Yagi T (1999) "Modulation of oxidative phosphorylation of human kidney 293 cells by transfection with the internal rotenone-insensitive NADH-quinone oxidoreductase (*NDI1*) gene of *Saccharomyces cerevisiae*" *Biochim. Biophys. Acta* **1412** 56-65

Boo Seo B, Nakamaru-Ogiso E, Flotte TR, Yagi T & Matsuno-Yagi A (2002) "A single-subunit NADH-quinone oxidoreductase renders resistance to mammalian nerve cells against complex I inhibition" *Mol. Therapy* **6(3)** 336-341

Buchanan SK (1999) "β-barrel proteins from bacterial outer membranes: structure, function and refolding" *Curr. Op. in Str. Biol.* **9** 455-461

Bykova NV, Rasmusson AG, Igamberdiev AU, Gardeström P & Møller IM (1999) "Two separate transhydrogenase activities are present in plant mitochondria" *Biochem. & Bioph. Res. Comm.* **265** 106-111

Carneiro P, Duaqrte M & Videira A (2004) "The main external alternative NAP(P)H dehydrogenase of *Neurospora crassa* mitochondria" *Biochim. Biophys. Acta* **160** 45-52

Carroll J, Fearnley IM, Shannon RJ, Hirst J & Walker JE (2003) "Analysis of the Subunit Composition of Complex I from Bovine Heart Mitochondria" *Mol. Cell. Proteomics* **2** 117-126

Carroll J, Fearnley IM, Skehel JM, Shannon RJ, Hirst J & Walker JE (2006) "Bovine Complex I Is a Complex of 45 Different Subunits" J. Biol. Chem. **281(43)** 32724-32727

Chaudhuri M, Ott RD, Saha L, Williams S & Hill GC (2005) "The trypanosome alternative oxidase exists as a monomer in *Trypanosoma brucei* mitochondria" *Parasit. Res.* **96(3)** 178-183

Chaudhuri M, Ott RD & Hill GC (2006) "Trypanosome alternative oxidase: from molecule to function" *Trends Parasitol.* **22(10)** 484-491

Chauveau M & Lance C (1991) "Purification and partial characterization of two soluble NAD(P)H dehydrogenases from *Arum maculatum* mitochondria" *Plant Physiol.* **98** 934-942

Clifton R, Millar AH & Whelan J (2006) "Alternative oxidases in *Arabidopsis*: a comparative analysis of differential expression in the gene family provides new insights into function of non-phosphorylating bypasses" *Biochim Biophys Acta* **1757(7)** 730-741

Cook-Johnson RJ, Zhang Q, Wiskich JT & Soole KL (1999) "The nuclear origin of the non-phosphorylating NADH dehydrogenases of plant mitochondria" *FEBS Lett.* **454** 37-41

Cottingham IR & Moore AL (1984) "Partial purification and properties of the external NADH dehydrogenase from cuckoo-pint (*Arum maculatum*) mitochondria" *Biochem. J.* **224(1)** 171-179

Crane EJ, Yeh JI, Luba J & Claibourne A (2000) "Analysis of the kinetic and redox properties of the NADH peroxidase R303M mutant: correlation with the crystal structure" *Biochemistry* **39(34)** 10353-10364.

Day IS, Reddy VS, Shad Ali G, & Reddy ASN (2002) "Analysis of EF-hand-containing proteins in *Arabidopsis" Genome Biology.* **3(10)** 0056.1–0056.24

DiMauro S & Schon EA (2003) "Mitochondrial respiratory-chain diseases" *N. Engl. J. Med.* **348** 2656- 2668

Duarte M, Peters M, Schulte U & Videira A (2003) "The internal alternative NADH dehydrogenase of *Neurospora crassa* mitochondria" *Biochem. J.* **371** 1005-1011

Dym O & Eisenberg D (2001) "Sequence-structure analysis of FAD-containing proteins" *Protein Science* **10** 1712-1728

Escobar MA, Franklin KA, Svensson ÅS, Salter MG, Whitelam GC & Rasmusson AG (2004) "Light regulation of the *Arabidopsis* respiratory chain. Multiple discrete photoreceptor responses contribute to induction of type II NAD(P)H dehydrogenase genes" *Plant Physiol.* **136** 2710-2721

Fang J & Beattie DS (2002a) "Novel FMN-containing rotenone-insensitive NADH dehydrogenase from *Trypanosoma brucei* mitochondria: isolation and characterisation" *Biochemistry* **41** 3065-3072

Fang J & Beattie DS (2002b) "Rotenone-insensitive NADH dehydrogenase is a potential source of superoxide in procyclic *Trypanosoma brucei* mitochondria" *Mol. Biochem. Parasit.* **123** 135-142

Fang J & Beattie DS (2003) "Identification of a gene encoding a 54 kDa alternative NADH dehydrogenase in *Trypanosoma brucei*" *Mol. Biochem. Parasit.* **127** 73-77

Finel M (1996) "Genetic inactivation of the H^+-translocating NADH:ubiquinone oxidoreductase of *Paracoccus denitrificans* is facilitated by insertion of the *ndh* gene from *Escherichia coli*" *FEBS Lett.* **393** 81-85

Fisher N & Rich PR (2000) "A motif for quinone binding sites in respiratory and photosynthetic systems" *J. Mol. Biol.* **296(4)** 1153-1162

Gomes CM, Bandeiras TM & Teixeira M (2001) "A new Type-II NADH dehydrogenase from the archeon *Acidianus ambivalens*: charaterisation and *in vitro* reconstitution of the respiratory chain" *J. Bioenerg. Biomembr.* **33(1)** 1-8

Grigorieff N (1998) "Three-dimensional structure of bovine NADH:ubiquinone oxidoreductase (complex I) at 22 å in ice" *J. Mol. Biol.* **277(5)** 1033-1046

Guénebaut V, Vincentelli R, Mills D, Weiss H & Leonard KR (1997) "Three-dimensional structure of NADH-dehydrogenase from *Neurospora crassa* by electron microscopy and conical tilt reconstruction" *J Mol. Biol.* **265(4)** 409-418

Guénebaut V, Schlitt A, Weiss H, Leonard KR& Friedrich T (1998) "Consistent structure between bacterial and mitochondrial NADH:ubiquinone oxidoreductase (complex I)" *J. Mol. Biol.* **276(1)** 105-112

Hinchliffe P & Sazanov LA (2005) "Organisation of iron-sulfur clusters in respiratory Complex I" *Science* **30** 771-774

Hinchliffe P, Carroll J & Sazanov LA (2006) "Identification of a Novel Subunit of Respiratory Complex I from *Thermus thermophilus*" *Biochemistry* **45** 4413-4420

Howitt CA, Udall PK & Vermass WFJ (1999) "Type 2 NADH dehydrogenases in the cyanobacterium *Synechocystis* sp. strain PCC 6803 are involved in regulation rather than respiration" *J. Bacteriology* **181(13)** 3999-4003.

Ikuma H & Bonner WD Jr. (1967) "Properties of higher plant mitochondria. I. Isolation and some characteristics of tightly-coupled mitochondria from dark-grown

mung bean hypocotyls" *Plant Physiol.* **42** 67–75

International Rice Genome Sequencing Project (2005) "The map-based sequence of the rice genome" *Nature* **436** 793-800

Jose Quiles M & Cuello J (1998) "Association of ferredoxin-NADP oxidoreductase with the chloroplastic pyridine nucleotide dehydrogenase complex in barley leaves" *Plant Physiol.* **117(1)** 235-44.

Joseph-Horne T, Hollomon DW & Wood PM (2001) "Fungal respiration: a fusion of standard and alternative components" *Biochim. Biophys. Acta* **1504** 179-105

Jaworowski A, Campbell HD, Poulis MI & Young IG (1981) "Genetic identification and purification of the respiratory NADH dehydrogenase of *Escherichia coli*" *Biochemistry* **20(7)** 2041-2047

Juszczuk IM & Rychter AM (2004) "Alternative oxidase in higher plants" *Acta Biochem. Polonica* **50(4)** 1257-1271

Kerscher SJ, Okun JG & Brandt U (1999) "A single external alternative NADH:ubiquinone oxidoreductase activity in *Yarrowia lipolytica*" *J. Cell Science* **112** 2347-2354

Kerscher SJ (2000) "Diversity and origin of alternative NADH:ubiquinone oxidoreductases" *Biochim. Biophys. Acta* **1459** 274-283

Kerscher SJ, Eschmann A, Okun PM & Brandt U (2001) "External alternative NADH:ubiquinone oxidoreductase redirected to the inner face of the mitochondrial inner membrane rescues complex I deficiency in *Yarrowia lipolytica*" *J. Cell Science* **114** 3915-3921

Kiefer H (2002) "In vitro folding of alpha-helical membrane proteins" *Biochim. Biophys. Acta* **1610** 57-62

Kitajima-Ihara T & Yagi T (1998) "Rotenone-intensive internal NADH-quinone oxidoreductase of *Saccharomyces cerevisiae* mitochondria: the enzyme expressed in *Escherichia coli* acts as a member of the respiratory chain in host cells" *FEBS Lett.* **421** 37-40

Knudten AF, Thelen JJ, Luethy MH & Elthon TE (1994) "Purification, characterisation and submitochondrial localization of the 32-kilodalton NADH dehydrogenase from Maize" *Plant Physiol.* **106** 1115-1122

Laemmli UK (1970) "Cleavage of structural proteins during the assembly of the head of bacteriophage T4" *Nature* **227** 680-685

Lemercier G, Bakalara N & Santarelli X "On-column refolding of an insoluble histidine tag recombinant exopolyphosphatase from *Trypanosoma brucei* overexpressed in *Escherichia coli*" *J. Chromatography* **768** 305-309

Lin S-J & Guarente L (2003) "Nicotinamide adenine dinucleotide, a metabolic regulator of transcription, longevity and disease" *Curr. Op. Cell Biol.* **15** 241-246

Luan S, Jörg Kudla I, Rodriguez-Concepcion M, Yalovsky S & Gruissem G (2002) "Calmodulins and Calcineurin B–like Proteins: Calcium Sensorsfor Specific Signal Response Coupling in Plants" *Plant Cell* **14** S389–S400

Luethy MH, Hayes MK & Elthon TE (1991) "Partial purification and characterization of three NAD(P)H dehydrogenases from *Beta vulgaris* mitochondria" *Plant Physiol* **97** 1317-1322

Luttik MAH, Overkamp KM, Kötter P, de Vries S, van Dijken JP & Pronk JT (1998) "The *Saccharomyces cerevisiea NDE1* and *NDE2* genes encode separate mitochondrial NADH dehydrogenases catalyzing the oxidation of cytosolic NADH" *J. Biol. Chem.* **273(38)** 24529-24534

Marres CA, de Vries S & Grivell LA (1991) "Isolation and inactivation of the nuclear gene encoding the rotenone-insensitive internal NADH:ubiquinone oxidoreductase of mitochondria from *Saccharomyces cerevisiea" Eur. J. Biochem.* **195(3)** 857-882

Marti-Renom MA, Stuart A, Fiser A, Sanchez R, Melo F & Sali A (2000) "Comparative protein structure modelling of genes and genomes." *Annu. Rev. Biophys. Biomol. Struct.* **29** 291-325.

Matsushita K, Otofuji A, Iwahashi M, Toyama H & Adachi O (2001) "NADH dehydrogenase of *Corynbacterium glutamicum.* Purification of an NADH dehydrogenase II homolog able to oxidize NADPH" *FEMS Microbiol. Letts.* **204(2)** 271-276

McDonald AE, Sieger SM & Vanlerberghe GC (2002) "Methods and approaches to study plant mitochondrial alternative oxidase" *Plant Physiol.* **116** 135–143.

Melo AMP, Roberts TH & Møller IM (1996) "Evidence for the presence of two rotenone-insensitive NAD(P)H dehydrogenases on the inner surface of the inner membrane of potato tuber mitochondria" *Biochim. Biophys. Acta* **1276** 133-139

Melo AMP, Duarte M & Videira A (1999) "Primary structure and characterisation of a 64 kDa NADH dehydrogenase from the inner membrane of *Neurospora crassa* mitochondria" *Biochim. Biophys. Acta* **1412** 282-287

Melo AMP, Duarte M, Møller IM, Prokisch H, Dolan PL, Pinto L, Nelson MA & Videira A (2001) "The external calcium-dependent NADPH dehydrogenase from *Neurospora crassa* mitochondria" *J. Biol. Chem.* **276(6)** 3947-3951

Melo AMP, Bandeiras TM & Teixeira M (2004) "New insights into Type II NAD(P)H:quinone oxidoreductases" *Microbiol. Mol. Biol. Rev.* **68(4)** 603-616

Menz RI & Day DA (1996a) "Identification and characterisation of an inducible NAD(P)H dehydrogenase from red beetroot mitochondria" *Plant Physiol.* **112** 607-613

Menz RI & Day DA (1996b) "Purification and characterisation of a 43-kDa rotenone-insensitive NADH dehydrogenase from plant mitochondria" *J. Biol. Chem.* **271(38)** 23117-23120

Michalecka AM, Svensson AS, Johansson FI, Agius SC, Johanson U, Brennicke A, Binder S & Rasmusson AG (2003): "Arabidopsis genes encoding mitochondrial type II NAD(P)H dehydrogenases have different evolutionary origin and show distinct responses to light" *Plant Physiology* **133** 642-652.

Michalecka AM, Agius SC, Møller IM & Rasmusson AG (2004) "Identification of a mitochondrial external NADPH dehydrogenase by overexpression in transgenic *Nicotiana sylvestris*" *Plant J.* **37** 415-425

Middelberg AJ (2002) "Preparative protein refolding" *Trends Biotech.* **20(10)** 437-443

Mitchell P (1961) "Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism" *Nature* **191** 144-148

Mitchell P & Moyle J (1965) "Stoichiometry of proton translocation through the respiratory chain and adenosine triphosphate systems of rat liver mitochondria" *Nature* **208**147-151

Møller IM, Johnston SP & Palmer JM (1981) "A specific role for Ca2+ in the oxidation of exogenous NADH by Jerusalem-artichoke (*Helianthus tuberosus*) mitochondria" *Biochem. J.* **194(2)** 487-495

Møller IM & Palmer JM (1982) "Direct evidence for the presence of a rotenone-resistant NADH dehydrogenase on the inner surface of the inner membrane of plant mitochondria". *Plant Physiol.* **54** 267–274.

Møller IM (2002)"A new dawn for plant mitochondrial NAD(P)H dehydrogenases" *Trends Plant Science* **7(6)** 235-237

Moore CS, Cook-Johnson RJ, Rudhe C, Whelan J, Day DA, Wiskich JT & Soole KL (2003) "Identification of AtNDI1, an internal non-phosphorylating NAD(P)H dehydrogenase in *Arabidopsis* mitochondria" *Plant Physiology* **133** 1968-1978

Nakayama S, Kawasaki H & Kretsinger R (2000): "Evolution of EF-hand proteins". In: *Calcium Homeostasis*. Edited by Carafoli E & Krebs J.: *Springer (NewYork)* 29-58.

Ojaimi J, Pan J, Santra S, Snell WJ & Schon EA (2002) "An algal nucleus-encoded subunit of mitochondrial ATP synthase rescues a defect in the analogous human mitochondrial-encoded subunit" *Mol. Biol. Cell* **13** 3863-3844

Pereira MM, Bandeiras TM, Fernandes AS, lemos RS, Melo AMP & Teixeira M (2004) "Respiratory chains from aerobic thermophilic prokaryotes" *J. Bioenerg. Biomem.* **36(1)** 93-105

Radermacher M, Ruiz T, Clason T, Benjamin S, Brandt U & Zickermann V "The three-dimensional structure of complex I from *Yarrowia lipolytica*: A highly dynamic enzyme" (2006) *J. Struct. Biol.* **154(3)** 269-279

RapisardaVA, Rodruíguez Montelongo L, Farias RN & Massa EM (1999) "Characterisation of an NADH-linked cupric reductase activity from the *Escherichia coli* respiratory chain" *Arch. Biochem. Biophys.* **370(2)** 145-150

Rapisarda VA, Chehín RN, De Las Rivas J, Rodríguez-Montelongo L, Farías RN & Massa EM (2002) "Evidence for Cu(I)-thiolate ligation and prediction of a putative copper-binding site in the *Escherichia coli* NADH dehydrogenase-2" *Arch. Biochem. Biophys.* **405** 87-94

Rasmusson AG, Svensson AS, Knoop V, Grohmann L & Brennicke A (1999) "Homologues of yeast and bacterial rotenone-insensitive NADH dehydrogenases in higher eukaryotes: two enzymes are present in potato mitochondria" *Plant J.* **20(1)** 79-87

Rasmusson AG & Agius SC (2001) "Rotenone-insensitive NAD(P)H dehydrogenases in plants: Immunodetection and distribution of native proteins in mitochondria" *Plant Physiol. Biochem.* **39** 1057-1066

Rasmusson AG, Soole KL & Ethlon TE (2004) "Alternative NAD(P)H dehydrogenases of plant mitochondria" *Ann. Rev. Plant Biol.* **55** 23-39

Rich PR (2003) "The molecular machinery of Keilin's respiratory chain" *Biochem. Soc. Trans.* **31(6)** 1095-1105

Roberts TH, Fredlund KM & Møller IM (1995) "Direct evidence for the presence of two external NAD(P)H dehydrogenases coupled to the electron transport chain in plant mitochondria" *FEBS Lett.* **373(3)** 307-309

Sambrook J, Fritsch EF & Maniatis T; Hrsg. (1989). "Molecular Cloning - A Laboratory Manual" 2nd Edition. *Cold Spring Harbour Laboratory Press*, New York.

Sazanov LA & Hinchliffe (2006) "Structure of the Hydrophilic Domain of Respiratory Complex I from *Thermus thermophilus*" *Science* **311** 1430-1436

Sazanov LA (2007) "Respiratory Complex I: Mechanistic and Structural Insights Provided by the Crystal Structure of Hydrophilic Domain" *Biochemistry* **46(9)** 2275-2287

Schmid R & Gerloff DL (2004) "Functional properties of the alternative NADH:ubiquinone oxidoreductase from *E. coli* through comparative 3-D modelling" *FEBS Lett.* **578(1-2)** 163-168.

Sherer TB, Betarbet R, Testa CM, Boo Seo B, Richardson JR, Kim JH, Miller GW, Yagi T, Matsuno-Yagi A & Greenamyre JT (2003) "Mechanisms of toxicity in rotenone models of Parkinson's disease" *J. Neuroscience* **23(43)** 10756-10764

Small WC & McAlister-Henn (1998) "Identification of a cytosolically directed NADH dehydrogenase in mitochondria of *Saccharomyces cerevisiae*" *J. Bacteriology* **180(16)** 4051-4055

Smith RH, Ding C & Kotin RM (2003) "Serum-free production and column purification of adeno-associated virus type 5" *J. Virol. Methods* **114** 115-124

Svensson ÅS & Rasmusson AG (2001) "Light-dependent gene expression for proteins in the respiratory chain of potato leaves" *Plant J.* **28(1)** 73-82

Svensson ÅS, Johansson FI, Møller IM & Rasmusson AG (2002) "Cold stress decreases the capacity for respiratory NADH oxidation in potato leaves" *FEBS Lett.* **517** 79-82

The Arabidopsis Genome Initiative (2000) "Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*" *Nature* **408** 798-815

Tuskan GA *et al.* (2006) "The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray)" *Science* **313(5793)** 1596-1604

Videira A & Duarte M (2002) "From NADH to ubiquinone in *Neurospora* mitochondria" *Biochim. Biophys. Acta* **1555** 187-191

Vincent P, Dieryck W, Maneta-Peyret L, Moreau P, Cassagne C, & Santarelli X (2004) "Chromatographic purification of an insoluble histidine tag recombinant Ykt6p SNARE from *Arabidopsis thaliana* over-expressed in *E. coli*" *J. Chromatography* **808(1)** 83-89

de Vries S, & Grivell LA (1988) "Purification and characterisation of a rotenone-insensitive NADH:Q6 oxidoreductase from mitochondria of *Saccharomyces cerevisiae*" *Eur. J. Biochem.* **176(2)** 377-384

de Vries S, van Witzenburg R, Grivell LA & Marres CAM (1992) "Primary structure and import pathway of the rotenone-insensitive NADH-ubiquinone oxidoresductase of mitochondria from *Saccharomyces cerevisiae*" *Eur. J. Biochem.* **203** 578-592

Wagner AM, Wagner MJ & Moore AL (1988) "*In Vivo* ubiquinone reduction levels during thermogenesis in *araceae*" *Plant Physiol.* **117** 1501-1506

Yagi T, Boo Seo B, Di Bernardo S, Nakamuaru-Ogiso E, Kao M-C & Matsuno-Yagi A (2001) "NADH dehydrogenases: From basic science to biomedicine" *J. Bioenerg. Biomem.* **33(3)** 233-242

Yagi T, Matsuno-Yagi A (2003) "The proton-translocating NADH-Quinone oxidoreductase in the respiratory chain: the secret unlocked" *Biochemistry* **42(8)** 2266-2274

Yeh JI, Claibourne A & Hol WG (1996) "Structure of the native cysteine-sulfenic acid redox center of enterococcal NADH peroxidase refined at 2.8 A resolution" *Biochemistry.* **35(31)** 9951-9957

Zolotukhin S, Potter M, Zolotukhin I, Sakai Y, Loiler S, Fraites Jr TJ, Chiodo VA, Phillipsberg T, Muzyczka N, Hauswirth WW, Flotte TR, Byrne BJ & Snyder RO (2002) "Production and purification of serotype 1, 2, and 5 recombinant adeno-associated viral vectors" *Methods* **28(2)** 158-167